



Db 307 IRHSFFDHGYNHLVDONFPLDPYWLSLLYKRLIGPKVLAVHVGAGLQKPRGVRIRDKL 366  
QY 421 RIYACTNNHHNNYVRSITLFIINLHRSRKKIKIAGLTKLVHQLVQYLLQPYGQGLKSK 480  
Db 367 RIYACTNNHHNNYVRSITLFIINLHRSRKKIKIAGLTKLVHQLVQYLLQPYGQGLKSK 426  
QY 481 SVOLNGQPLVMYDDGTLPELKPRLRAGRTLPIPVMTGFFVVKVNNALACRYR 534  
Db 427 SVOLNGQPLVMYDDGTLPELKPRLRAGRTLPIPVMTGFFVVKVNNALACRYR 480

RESULT 2  
T45608  
hypothetical protein F13G24.30 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T45608  
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba  
submitted to the Protein Sequence Database, December 1999  
A:Reference number: 223009  
A:Accession: T45608  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-521 <BEV>  
A:Cross-references: EMBL:AL133421  
A:Experimental source: cultivar Columbia; BAC clone F13G24  
C:Genetics:  
A:Map position: 5  
A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3  
A:Note: F13G24.30

Query Match 10.8%; Score 302; DB 2; Length 521;  
Best Local Similarity 24.8%; Pred. No. 2e-15;  
Matches 139; Conservative 73; Mismatches 183; Indels 166; Gaps 27;

QY 75 NENFLSLQID--PSIHD-----GWLDFLS---SKRLVTLA-RGLSPAFRLFGGR 119  
Db 22 DENFVCATLDWNP---HDKACNYDCPWGYSSVINMDLTRPLTKAIKAFKPLRIRIGSL 78  
QY 120 TDFLQFQ--NLNRP-----KSRGGPGPDYLLKNEYDDIVRSVDALDKQKCKIAQHPVM 173  
Db 79 QDOVIYDVGNLKTCPRPFOKMNSG-----LFGFSKGCGLHMKRW-- 117  
QY 174 LELOREKAAQMHVL-----LLKEQFSNTYSNLILTEPNRYRMHGRAVN--- 217  
Db 118 -ELNSFLTATGAVVTGLNALRGRHKLRKANGGAWDHINTQDFLNTYVSKGYVIDSWEF 176  
QY 218 GSQI-----GKDYIQLKSLQPIRIYSRASLYGPNIGRPRKNVIALDGF- 262  
Db 177 GNELSGSGVGASVSAELYCKDLIVLKDVIN--KVYKNSWLHKPILVAP-----GGFY 226  
QY 263 -----MKVAG-STVDVATWQHCVIDGR-----VVKVMDFLKTRLDLTSDQIRKIQ 307  
Db 227 EQOQWYTKLEISGPSYVDVVT-HHIYNLGSNDPALVKKIMD---PSYLSQVSKTFKQYN 282  
QY 308 KVVNTVTPGKKIWLEGWTTTSAG-----GTNNLSDSVAAGFLWLNTLGLMANQIDVVIH 363  
Db 283 QTQEHGP-----WASFWGESGAYNSGGRHVSDFIDFSWLDQLGMSARHNTKYVCQ 338  
QY 364 SFFDHGYNHLVDONFPLDPYWLSLLYKRLIGPKVLAVHVGAGLQKPRGVRIRDKLRY 423  
Db 339 TLVGGFYGLLEKTFVPNPDPYSALLHRLMGKGLAVQTDG-----PPQLRYV 387  
QY 424 AHCNTNNHHNNYVRSITLFIINLH-----RSBKKIKLACTLRDKLV 464  
Db 388 AHCCKG-----RAGVTLLINLSNQSDFTVSVNSGINVVLNAESRKKSLDITLRKPS 441  
QY 465 -----HOYLLQPYGQEG-LKSKSVOLNGQPLVMYDDGTLPELKPRLRAGRT 510  
Db 442 WIGSKASDGYLNREEYHLTP--ENGVLRSKTYWLNKSLKPTATGDIPLSVLRVSNVP 499  
QY 511 LVTPPVTMGFFVVKVNNALAC 531

Db 500 LNVLPLSMSFVLPNFDASAC 520

RESULT 3  
T01953  
hypothetical protein T2L5.6 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 21-Jan-2000  
C:Accession: T01953  
R:Geisel, C.; Smith, A.; Le, T.  
submitted to the EMBL Data Library, October 1998  
A:Description: The sequence of A. thaliana T2L5.  
A:Reference number: Z14470  
A:Accession: T01953  
A:Status: translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-190 <GEI>  
A:Cross-references: EMBL:AF096371; NID:g3695386; PID:g3695392  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4  
A:Introns: 36/2; 69/3  
A:Note: T2L5.6  
C:Superfamily: Arabidopsis thaliana hypothetical protein T2L5.6

Query Match 5.6%; Score 155; DB 2; Length 190;  
Best Local Similarity 24.9%; Pred. No. 0.00011;  
Matches 49; Conservative 32; Mismatches 70; Indels 46; Gaps 6;

QY 362 RHSEFFDHGYNHLVDONFPLDPYWLSLLYKRLIGPKVLAVHVGAGLQKPRGVRIRDKL 421  
Db 12 RQSLGGNGYGLLNTNFTPNPDYYSALLWRQLMGRKALFTTFSGTK-----KIR 60  
QY 422 IYAHCTNNHHNNYVRSITLFIINLHRSR---KTKLAGTLRDKLVHQLVQYLLQPY----- 472  
Db 61 SYTHCAROSK-----GITVLLMLNDLNTTVVAKVELNNSF--SLRHTKHKMSYKRASSQ 112  
QY 473 ---GOEG-----LKSKSVOLNGQPLVMYDDGTLPELKPRLRAGRTLVP 514  
Db 113 LFGPNGVIOREYHLTKAGDNLHQSQTMLLNGNALQVNSMGDLPPIEPHINSTEPIIA 172  
QY 515 PVTMGFFVVKVNNALAC 531  
Db 173 PYSIVFVHMNRNVVPAC 189

RESULT 4  
T14022  
reverse transcriptase homolog - slime mold (Dictyostelium discoideum) retrotransposab  
C:Species: Dictyostelium discoideum  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 11-May-2000  
C:Accession: T14022  
R:Winckler, T.; Tschepke, C.; de Hostos, E.L.; Jendretzke, A.; Dingermann, T.  
Mol. Gen. Genet. 257, 655-661, 1998  
A:Title: Tdd-3, a transfer RNA gene-associated poly(A) retrotransposon from Dictyoste  
A:Reference number: Z17858; MUID:98265925  
A:Accession: T14022  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-1260 <WTN>  
A:Cross-references: EMBL:AF002669; NID:g2558820; PID:g2558822; PIDN:AAC48324.1

Query Match 4.0%; Score 111; DB 2; Length 1260;  
Best Local Similarity 19.7%; Pred. No. 5.2; Mismatches 89; Indels 232; Gaps 34;

QY 69 NPVRTVNEFLSLQIDPPSIHDGWL---DFL---SKRLV-----TLARGLSP 110  
Db 94 NGIGILNHNQNKLSPIIEGRLLISDILIKDTTTRILAIYAPAPQDKRKTASTLNK 153

F98229  
sarcosine oxidase alpha chain PA5418 [imported] - Agrobacterium tumefaciens (strain C58,  
C:Species: Agrobacterium tumefaciens  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C:Accession: F98229  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; PMID:11743194  
A:Accession: F98229  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-986 <KUR>  
A:Cross-references: GB:AE007870; PID:gl5159208; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L\_1574  
A:Map position: linear chromosome

6  
RESULT  
AH3056  
sarcosine oxidase alpha subunit [imported] - Agrobacterium tumefaciens (strain C58, D  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C:Accession: AH3056  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCI  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Ka  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AH3056  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-986 <KUR>  
A:Cross-references: GB:AF008689; PIDN:AA144870.1; PID:gl17742518; GSPDB:GN00187  
A:Experimental source: strain C58 (Dupont)

[illegible]

Db 362 DQRLDEARALGIEVLAGHSVYTAGRLRVSSMTVGRNGSGNKRKIAIDALVVSAGW--- 418  
QY 96 LSSKRLVTLARGLSPAFLEGGKRTDFLOFQNLRN-----PAKSRGG 137  
Db 419 TSPVHLFSQSRGK----LKFDAAQNRPLPDHIVQNCVSGACGNTDLDVAITAEAAAGG 474  
QY 138 PGDPYILKN-----YEDDIVRSDVALDKQKQKIAQHFD 171  
Db 475 GSAFSGENARAWTGMGIAAGAGGTGVKAFIDFQHDVCAKDIRLAVREG----- 526  
QY 172 VMLELQREKAAQMHVLLKLEQFNTSYNLLITPEPNRYRWHGRAVNSOLGKDY--IQLK 229  
Db 527 -----MHSV---EHIKRFTTNGMASDQGRKSMHGLAIASEALGRDLPRVGLT 571  
QY 230 SLLQP-----IRIYSRASLYGPNIGRPKNVIALLDG-----EMKVAGS 268  
Db 572 TFRQPTPTVFTGLINHSRGALFDPRTKTPMHEE-ELAAGAVFEDVGNKRAWFFPRAGE 530  
QY 269 TVDAVTWQHICYIDGRVVKVNDFLKTRLDLSDQIRKIQKV-----NTYTPGKKIWLLEG 323  
Db 631 DMHEAINRECKTVRTSGVGFEDASTLGIKIEVVGPDAAKFLNLIVTNAWDTLKPGRCRY--G 688  
QY 324 VVTSAGTNLSDSYAAGFLWLNTLGLMLANOGIDVVIHRSFEDHCYNHLVDQNFPLPD 383  
Db 689 IMTREGFV-----YDDG-----VVGRLAEDRFHTTTTGAPRVLOHMDYLQTEFPD 737  
QY 384 --YWLSLLYKR-----LIGPK--VLAVHVAGLQKRP-----PGRV--- 415  
Db 738 LNVWLTSATEQMAVIAVQPKAREVIAPFVEGIDLSPEAFPHMAVAEGKFCGVPRFLRV 797  
QY 416 -----IRDKLRIYACHTNHHNHNHNVVRSITLFIINLH----- 447  
Db 798 SFTGELGFEINVPADYGAAVSAIRDTEAVGCG-----LYGTETMHLRAEKGYII 849  
QY 448 -----RSRKKIKLAGTLRDKLVHVOYLLQPYGOEGL 477  
Db 850 VGQDGTGTVPDAGLAWAVSKKTKDFVGIKRLIDLTTRTKQKLV-----GL 898  
QY 478 KKSQVOLNQPLVMVDG-----TLPELKPRLP----- 505  
Db 899 KTK-----DRLTVPDEGGQIVTDPN-QPKPMTMLGHVTSAYWSENGLHGSIAFALVADGR 951  
QY 506 -RAGRTLVP 514  
Db 952 ARMGETLYIP 961  
RESULT 7  
OVBX  
adenylate cyclase (EC 4.6.1.1) - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: ATP pyrophosphate-lyase; protein J1401; protein YJL005w  
C:Species: *Saccharomyces cerevisiae*  
C>Date: 28-Dec-1987 #sequence\_revision 08-Sep-1995 #text\_change 21-Jan-2000  
C:Accession: S56776; S56775; A24776; S05828; S55183  
R:To Van, D.; Perea, J.; Jacq, C.  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56776  
A:Accession: S56776  
A:Molecule type: DNA  
A:Residues: 1-1823 <DEB>  
A:Cross-references: EMBL:Z49280; GSPDB:GNO0010; MIPS:YJL005w  
R:de Haan, M.; Grivell, L.A.; Smits, P.H.M.  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56771  
A:Accession: S56775  
A:Molecule type: DNA  
A:Residues: 673-2026 <ZAG>  
A:Cross-references: EMBL:Z49280; GSPDB:GNO0010; MIPS:YJL005w  
R:Kataoka, T.; Broek, D.; Wagler, M.  
Cell 43, 493-505, 1985  
A:Title: DNA sequence and characterization of the *S. cerevisiae* gene encoding adenylate  
A:Reference number: A24776; MUID:86079531  
A:Accession: A24776

A:Molecule type: DNA  
A:Residues: 1-261, 'L', 263-547, 'L', 549-591, 'H', 593-708, 'I', 710-961, 'P', 963-1387, 'S', 13  
A:Cross-references: EMBL:M12057; NID:g171359; PIDN:AAA34549.1; PID:g171360  
A>Note: the authors translated the codon TTA for residue 262 as Ser, ACG for residue  
2 as Leu, TCA for residue 1566 as Ala, GAG for residue 1659 as Ala, GGT for residue 1  
R:Masson, P.; Lenzen, G.; Jacquemin, J.M.; Danchin, A.  
Curr. Genet. 10, 343-352, 1986  
A:Title: Yeast adenylate cyclase catalytic domain is carboxy terminal.  
A:Reference number: S05828; MUID:88165073  
A:Accession: S05828  
A:Molecule type: DNA  
A:Residues: 1042-1426, 'D', 1428-1460, 'T', 1462-1955, 'V', 1957-2008, 'M', 2014-2015, 'TNEFKV  
A:Cross-references: EMBL:X03449; NID:g3487; PIDN:CAA27175.1; PID:g3488  
R:de Haan, M.; Smits, P.H.M.; Grivell, L.A.  
submitted to the EMBL Data Library, May 1995  
A:Reference number: S55183  
A:Accession: S55183  
A:Molecule type: DNA  
A:Residues: 673-2026 <DEW>  
A:Cross-references: EMBL:X87611; NID:g854567; PIDN:CAA60917.1; PID:g854568  
C:Genetics: SGD:CYR1; CDC35; MIPS:YJL005w  
A:Gene: SGD:CYR1; CDC35; MIPS:YJL005w  
A:Cross-references: SGD:S0003542; MIPS:YJL005w  
A:Map position: 10L  
C:Superfamily: yeast adenylate cyclase; leucine-rich alpha-2-glycoprotein repeat homo  
C:Keywords: CAMP biosynthesis; duplication; phosphorus-oxygen lyase; tandem repeat  
F:669-1343/Region: leucine-rich 23-residue repeats  
F:1085-1087/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR>  
F:1610-2026/Domain: catalytic #status predicted <CAT>  
F:1664-1749/Domain: yeast adenylate cyclase catalytic domain homology <YACC>

Query Match 3.8%; Score 107; DB 1; Length 2026;  
Best Local Similarity 20.2%; Pred. No. 22;  
Matches 124; Conservative 82; Mismatches 207; Indels 200; Gaps 31;  
QY 37 LSSQAGRRPLPVDRAAGLKE-KTLILLDVSTKNPV---RTVNFNLSLQLDPST----- 87  
Db 891 LNLQCNELSLP---AGFVELKLNQLLDLSSNKPMHPYEVINYCTNLLQIDLNSYKIQS 946  
QY 88 -----IHGWLDFLSKKRLVTLARGLSAFLEFGCKRTDFLOFQNL----- 128  
Db 947 LPQSTKYLVKLAKMNLHKNLNFIGDLSEMTDLRLTLNLRNRISSIKTNASNLQNLFLTD 1006  
QY 129 -----RNPAGSRGGPGDYIYLNKVEDDITVRSDDVALDKQKCKIAQH 169  
Db 1007 NRISNFDTLPKLRALEIQENPITSIS--FKDFYPKN-----MTSLTLNK---AQLSSI 1055

QY 170 PDVMLE----LOREKAAQMHVLLKLEQFNTSYNLLITPEPN-----YRT- 210  
Db 1056 PGEILLTKLSFLEKLELNQNNLTRLPQELSKTLKLVFLSVARNKLEYIPPELSQLSLRTL 1115  
QY 211 -MHGRAVNSQLGKDYIOLKSLQPIRIYSRASL-----YCPNIGRPKNVIALLD 260  
Db 1116 DLHNSNIRDFVDGMENLELTSLNISSNAFNGSSLENSFYHNMYSYKLSK-----SLM- 1168  
QY 261 GFMKVAGSTVDVATWQ--HCYIDGRVVKVMDFLKTRLDLSDQIRKIQKVNTYTPGKK 318  
Db 1169 -FFTAADNQFDAMWPLENCFVN---LKVNLNYSNFDV---SHMKLESTELYLSNKK 1221  
QY 319 IWLGEVVTTSAGGTNNLSDSYAAGFLW--LNTLGLMANQGDV-----VIHRSFFDHGY 370  
Db 1222 L-----TTLSGDT-----VLKWSLTKTLMNSNQMLSLPAELNSLSQLSVDFVGA 1266  
QY 371 NHLVDQNFNPLPDY-WLS---LLYKRLIGP-----KVLA 400  
Db 1267 NQKYNISNYHDWNWRNKKELKYNLNSGNRRFEIKSFISHDIDADLSLTVLPOLKVLG 1326  
QY 401 VHVAGLQKRPGRVIRDKLRIYACHTNHHNH-----NYV-----RGSTTL 441  
Db 1327 LMDVTLNTTKVPDENVNFRLRTTASIINGMYGVADTLGQRDYVSSRDVTFERFRGNDE 1386  
QY 442 FIINLHRSRKK-----IKLAGTLRDK-LVHQVLLQPYGQE-----GLKSKSVOLN 485







Query Match	3.6%	Score 100.5;	DB 2;	Length 582;
Best Local Similarity	21.2%	Pred. No. 10;		
Matches	69;	Conservative	49;	Mismatches 110; Indels 97; Gaps 14;

  

QY	49	VDRAGLKEKTLILDDYSTKNPVTNENFLSLQDPSIIHDGWLDFLSKRLVTLARGL	108
DB	58	VQVQSVVDRILIIAITE---RTVDENDELHADSIILYD-----	95
QY	109	SPAFLEFGGKRTDFLQFNLRNPASRGPGPDYILKNYEDD---IVRSDVALDKQKC	164
DB	96	-----GDRTEAFDE---SPAEV---TGLDIPKDLAEDAGGAIMRNIVALGAV--C	139
QY	165	KIAQHPDVMLEQREKAAQHMLVLLKQFSNTYSNLILTEPNVYRTWMHGRAVNSQLGKD	224
DB	140	AVADFPFIENLDESLEK-----RFSKGGEQII---ENNKQ-----AARLGAE	177
QY	225	YI---OLASLLQPIRIYSRASLYGPNIGRPKNVITALLDGFPMKVGASTVDATWQHCVIDG	282
DB	178	YVAEFEDVTLPYELETTDEDYVLLNGDEAIGMGAIAGCRFYAGYDITPAT-----	229
QY	283	RVKVMDFLKTRL-----LDTLSDQIRKIQKVNTYTPGKKIKWLGEVVTTSAGGTNNLS	336
DB	230	--DVMEYLTGRISQFGHHVQAEDELAANLALGAARAGAR---SMTATSGPGIDLMS	282
QY	337	DSYAAGFLWLNTGLMANOGIDVVI	361
DB	283	E-----TFGLVATSETPLVI	297

  

RESULT	15
C71374	

probable glucose inhibited division protein A (gida) - syphilis spirochete  
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 26-Aug-1999  
C:Accession: C71374  
R:Fraser, J.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G.  
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M.  
rson, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998  
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A:Reference number: A71250; MUID:98332770  
A:Accession: C71374  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-630 <COL>  
A:Cross-references: GB:AE001189; GB:AE000520; NID:G3322293; PIDN:AAC65038.1; PID:G332  
A:Experimental source: strain Nichols  
C:Genetics:  
A:Gene: TP0044  
C:Superfamily: gida protein

  

Query Match	3.6%	Score 100.5;	DB 2;	Length 630;
Best Local Similarity	21.3%	Pred. No. 12;		
Matches	77;	Conservative	52;	Mismatches 129; Indels 103; Gaps 20;

  

QY	218	GSQLGK--DYIQLKSLQPIRIYSRASLYGPNIGRPKNVIALDGFPMKVGASTVDATW	275
DB	68	GGEMGFADACMIQ-----YRLLNKSR--GPAVQAPRIQADKFL--YAAQVKVTTLECTQH	118
QY	276	QHCYLDGRVVKVMDFLKLRLDLSQIRKIQKVNTYTPGKKIKWLGEVVTTSAGGTNNL	335
DB	119	LHLQD-----TVDVVCSTNTDAGYVAGAAHVVT--ARGRRISARAVVLTG-----	166
QY	336	SDSYAAGFLWL-----NTLGLMANOGIDVVIHRSFFDHG-----YNHLVDQNF	378
DB	167	--TEMEGRVYIGEYAEPEGRLEGEHAAEGLAALKKGFQMGRLKTGTTPARVLRKSV--	221
QY	379	NPLFDYWLSLLYKR-----LIGPKVLAVHVAGLQKRKPRGVRIDKLRIYAHCTNNHHNY	434
DB	222	-----LSVMEKEQADAMPFSSFA--HV-----EINRPHADCYINYNTERHQ	263

Qy 435 VRGSTITFIINLHRS---RKKIKLAGT-----LRDKLVHQYLLQPYGOGLKS 479  
Db 264 IRE-----NFHRSPFTSGRIKAVGTRYCPSIEDKVRKFPDRIRHQLYIEP---EGLDT 313  
Qy 480 KSVQLNGQPLVMVDD-----GTLPELK-----PRDLRAGRTLVIPIPPVTMGF-FVVKNVNA 528  
Db 314 EELYINGLSSCLPEDIQDEMINTIPGMERAVITRPAYAVDYAVLFPVQLGIDLQTKRVSG 373  
Qy 529 L 529  
Db 374 L 374

Search completed: July 30, 2002, 08:16:22  
Job time: 339 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:15:13 ; Search time 36.14 Seconds  
(without alignments)  
572.115 Million cell updates/sec

Title: US-09-836-461-2

Perfect score: 2785

Sequence: 1 MRVLCAPPEMPSSNSRPPA.....PVTMGFFVKNVNALACRYR 534

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	3.8	2026	1	CYAA_YEAST
2	101.5	3.6	629	1	T3MH_HAEIN
3	100.5	3.6	630	1	GDA_TREPA
4	98.5	3.5	335	1	SWA_BUCAI
5	97.5	3.5	845	1	CLPC_CHLPN
6	97.5	3.5	1517	1	RPOC_CAMJE
7	96.5	3.5	969	1	DP3A_UREPA
8	96	3.4	1229	1	SLP3_YEAST
9	96	3.4	4367	1	DHHC_NEUCR
10	94.5	3.4	682	1	SNK_RAT
11	94.5	3.4	685	1	SNK_HUMAN
12	94.5	3.4	690	1	HELS_METH
13	94.5	3.4	1054	1	POL_STVMK
14	94	3.4	625	1	GIDA_LACLA
15	93.5	3.4	412	1	SYI2_BACSU
16	93.5	3.4	947	1	LKTA_PASSP
17	93	3.3	488	1	KI15_CAEEL
18	93	3.3	573	1	DPOL_MOUSE
19	93	3.3	849	1	PHSG_SYNY3
20	92.5	3.3	682	1	SNK_MOUSE
21	92.5	3.3	1616	1	VITL_CAEEL
22	92.5	3.3	2241	1	TEGU_HCMVA
23	92	3.3	461	1	GP13_YEAST
24	92	3.3	534	1	COX1_KLULA
25	92	3.3	535	1	COX1_HANWI
26	92	3.3	824	1	TC37_HUMAN
27	92	3.3	898	1	ACOC_CUCMA
28	91.5	3.3	536	1	DIT1_YEAST
29	91.5	3.3	1259	1	LN1_HUMAN
30	91	3.3	498	1	C72U_ARATH
31	91	3.3	676	1	YO7T_YEAST
32	91	3.3	1032	1	WT18_YEAST
33	90	3.2	948	1	RPOD_PODAN

ALIGNMENTS

RESULT 1  
CYAA\_YEAST

ID	CYAA_YEAST	STANDARD;	PRT; 2026 AA.
AC	P08678;		
DT	01-JAN-1988 (Rel. 06, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Adenylate cyclase (EC 4.6.1.1) (APP pyrophosphate-lyase) (Adenylate cyclase).		
DE	CYRI OR CDC35 OR HSRI OR SRA4 OR YJL005W OR J1401.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
OX	NCBI_TaxID=4932;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=86079531; PubMed=2934138;		
RA	Kataoka T., Broek D., Wigler M.;		
RT	"DNA sequence and characterization of the S. cerevisiae gene encoding adenylate cyclase."		
RL	Cell 43:493-505(1985).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	To Van D., Perea J., Jacq C.;		
RL	Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=S288C / FY1679;		
RA	de Haan M., Smits P.H.M., Grivell L.A.;		
RL	Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.		
RN	[4]		
RP	SEQUENCE OF 1042-2026 FROM N.A.		
RX	MEDLINE=88165073; PubMed=3327602;		
RA	Masson P., Lenzen G., Jacquemin J.M., Danchin A.;		
RT	"Yeast adenylate cyclase catalytic domain is carboxy terminal."		
RL	Curr. Genet. 10:343-352(1986).		
RN	[5]		
RP	MUTAGENESIS OF THR-1651.		
RX	MEDLINE=91122042; PubMed=1991451;		
RA	Feger G., de Vendittis E., Vitelli A., Masturzo P., Zahn R.,		
RA	Verrotti A.C., Kavounis C., Pal G.P., Fasano O.;		
RT	"Identification of regulatory residues of the yeast adenylate cyclase."		
RL	EMBO J. 10:349-359(1991).		
CC	!- FUNCTION: PLAYS ESSENTIAL ROLES IN REGULATION OF CELLULAR METABOLISM BY CATALYSING THE SYNTHESIS OF A SECOND MESSENGER, CAMP.		
CC	!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.		
CC	!- ENZYME REGULATION: THE PRESENCE OF GTP-BOUND RAS2 PROTEIN IS REQUIRED IN ORDER TO ELICIT A MAGNESIUM-DEPENDENT ADENYLATE CYCLASE ACTIVITY.		
CC	!- SIMILARITY: BELONGS TO ADENYLATE CYCLASE CLASS-3 FAMILY.		
CC	!- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).		
CC	!- SIMILARITY: CONTAINS A PP2C-LIKE DOMAIN.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		

Q06807 bos taurus  
P11369 mus musculus  
P07702 saccharomyc  
P42787 drosophila  
P53760 gallus gall  
P57844 pasteurella  
Q93179 epstein-bar  
Q937m0 chlamydia p  
Q49937 spinacia ol  
O10783 mycobacteri  
Q12979 homo sapien  
P14376 escherichia



```

RX MEDLINE=98032770; PubMed=9655876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RA "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -1- FUNCTION: NOT KNOWN.
CC -1- SIMILARITY: BELONGS TO THE GIDA FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001189; AAC65038.1; -
CC TIGR; TP0044; -
CC InterPro: IPR002218; GIDA.
CC Pfam: PF01134; GIDA; 1.
CC ProDom: PD003738; GIDA; 1.
CC PROSITE; PS01280; GIDA_1; 1.
CC PROSITE; PS01281; GIDA_2; 1.
CC Complete proteome.
CC SEQUENCE 630 AA; 70175 MW; 1B52C6F02C1EC275 CRC64;
SO

```

Query Match			3.6%; Score 100.5; DB 1; Length 630;
Best Local Similarity			21.3%; Pred. No. 4.5;
Matches	77; Conservative	52; Mismatches	129; Indels 103; Gaps 20;
Qy	218	GSQLGK--DYIQLKSLLQPIRIYSRASLYGNIGRPKNVTALLDGFNMKVAGSTVDATW	275
Db	68	GGEKGKFADACMIQ-----YRLLNKSR--GPAVQAPRIQADKFL--YAQVKYTLECTQH	118
Qy	276	QHCVIDGRVVVMDFLKTRLDLTSLDSQIRKIQKVNTTYTPGKKTWLEGVTVTSAGGTNNL	335
Db	119	LHLHYOD-----TWDDVCSTNTDAGVAYGAHAHVTT--ARGRRISARAVLTGT-----	166
Qy	336	SDSYAAGFLWL-----NTLCMLANOGIDVVIRHSFFDHG-----YNHLVDQNF	378
Db	167	--TWEGRVYIGEYEAPGRGLGEHAAEGLGAALRKKGGFOMGLTKGTGPARVLRSVD---	221
Qy	379	NPLPDYWLSLYKR----LIGPKVLAVHAVGLQRKPGRVIRDKLRTIYAHCTHHNHNY	434
Db	222	-----LSVMKEQEAADAIIMRFESFA-HV-----EINRHADCYIVNTERTHQL	263
Qy	435	VRGSITLEFIINLHRS---RKIKLAGT-----LRDKLVHQVYLLOPYGOEGLKS	479
Db	264	IRE-----NFHRSPFFSGRIKAVGTGYCPSIEDKVRKFPDRIRHQLYIEP---EGLDT	313
Qy	480	KSVOLNGQPLYMVD-----CTLPELK----PRPLRAGRITVIPPVMTGF-FVVKVNNA	528
Db	314	EELYINGLUSSCLPEDIQDEMIRTIPGMERAVITRPAYADVAVLPVQGLDLOTKRRVSG	373
Qy	529	L 529	
Db	374	L 374	
RESULT 4			
SYW_BUCAI	ID	SYW_BUCAI	STANDARD; PRT; 335 AA.
AC	P57602;		
DT	16-OCT-2001	(Rel. 40, Created)	
DT	16-OCT-2001	(Rel. 40, Last sequence update)	
DT	16-OCT-2001	(Rel. 40, Last annotation update)	

```

DE DE (TrpRS).
GN TRPS OR BU536.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrtosiphon pisum
OC symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RL Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) = AMP +
CC diphosphate + L-tryptophanyl-tRNA(Trp).
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP001119; BABI3229.1; -.
DR InterPro; IPR001412; trna-synt_I.
DR InterPro; IPR002306; trna-synt_trp.
DR Pfam; PF00579; trna-synt_1b; 1.
DR PRINTS; PR01039; TRNASYNHTRP.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 14 22 "HIGH" REGION.
FT SITE 196 200 "KMSKS" REGION.
FT BINDING 199 199 ATP (BY SIMILARITY).
FT SEQUENCE 335 AA; 38549 MW; 871562D5A6734E3F CRC64;
-----
Query Match 3.5%; Score 98.5; DB 1; Length 335;
Best Local Similarity 21.1%; Pred. No. 2.6;
Matches 50; Conservative 38; Mismatches 78; Indels 71; Gaps 10;
QY 151 IVRSVDALDKQCKIAQHPDVMLELQREKAAQMHVLVLLKEQFSNTYSNLIITERNNYT 210
DB 132 LMAADILLYQTNPVPVQDQKHVELTRN-----IAHRENSLYGH-VETLPKPLIT 181
QY 211 MHGRAVNSQLGKDYIQLKSLLOPIRIYSRASLYGNIGRPKNVYALDGG---FMKVA 266
DB 182 QHGS-----KIMSLEPSKMSKSDI-----NKNVIFLLDDDKTWTISKIQ 222
QY 267 GSTVDAVTWQHCVID-----GRVVKVM-----DFL-----KTRLLDPL 299
DB 223 NAYTSETPSKIYDEKPKPGISNLEILSALTNRKDIDILLKEGMYSEFNIVADHL 282
QY 300 SDOIIRKIQVNTYTPG-----KKIWLEGVVTTSAGTNNLSDSYAAGFWLNTLGL 352
DB 283 SKFLYKLOKSYNDRNDEVYLLAKIAVEGAMKSQLSKNLTTRKY-----DKLGLI 332
-----
RESULT 5
CLPC_CHLPN
ID CLPC_CHLPN STANDARD; PRT; 845 AA.
AC Q9Z8A6; Q9K297; Q9JSB9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable ATP-dependent Clp protease ATP-binding subunit.
GN CLPC OR CPN0437 OR CP0316.

```

```

OS Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heldelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linhorst K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -!- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY. CLPC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001626; AAD18581.1; -.
DR EMBL; AE002194; AAF38172.1; -.
DR PHCI-2DPAGE; Q9Z8A6; -.
DR TIGR; CP0316; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR003959; AAA_subfam.
DR InterPro; IPR001270; CLP_AB.
DR InterPro; IPR004176; Clp_N.
DR InterPro; IPR001943; UVR.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF02861; Clp_N; 2.
DR Pfam; PF02151; UVR; 1.
DR PRINTS; PR00300; CLPPROTEASEA.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00870; CLPAB_1; 1.
KW Chaperone; ATP-binding; Repeat; Complete proteome.
FT NP_BIND 232 239 ATP (POTENTIAL).
FT NP_BIND 569 576 ATP (POTENTIAL).
FT CONFLICT 4 4 K -> T (IN REF. 3).
FT CONFLICT 141 141 K -> R (IN REF. 1).
FT SEQUENCE 845 AA; 94980 MW; C9D9205458812866 CRC64;
-----
Query Match 3.5%; Score 98; DB 1; Length 845;
Best Local Similarity 19.8%; Pred. No. 11;
Matches 131; Conservative 89; Mismatches 241; Indels 202; Gaps 30;
QY 13 SSNSRPPACLPAGALYALALLHLSSSQAGDR-----RLPVD----RAA 53
DB 160 SSRSNPFSSKSP-----LGHSLGSDRNEKLSALKAYGYDLTEMVRESKLDPPVIGRSS 211

```



QY 54 GLKEXTLLDVSSTKPNV-----RTVNENFLSLQDPSIIHDGWLDFLSKRLVTTIARGLS 109  
 Db 212 EVERILILCLRRRKNPNVLIGEAGVGKTAIVEGLAQKIIILAEVDPALRKRRLITLDLALM 271  
 QY 110 PAFIRFGG-----KRTDFLOF-----ONLRNPAKSR- 135  
 Db 272 IAGTKYRQOFERIKAVMDYERKHCNILLFIDELHTIVGAGAAEADASNIIKLPAALARG 331  
 QY 136 -----GGPGDYKKNYEDDIRSDVALDKQCKIAQHP-----DVMELQR-----EKAQOM 184  
 Db 332 EIQIGATTIDEYRKHIE-----KDAALER-RFQIVVHPSPVDETIBILRGLKKKYEHEH 385  
 QY 185 HLVLKQFQSNYSNLIITENPNYTMGR-----AVNSQLQKDIQ 227  
 Db 386 HNVFITEEALKAAATL-----SDQYVHGRFLPKDAIDLLDEAGARVRVNTMGQPTDLMK 439  
 QY 228 LKSLQPIRIYRSASLYGPNIGRPNKKNVIALLDGPMKVAGSTVDVAV--TWQH-----CY 279  
 Db 440 LEAEIENTKLAKEQA-----IGTQYEKAAGLRDEKKLRERLQSMKQEWENHKEHQVP 494  
 QY 280 IDGRVVKVMDFLKT-----RLDLTSLDQIRKIQKVNTYTPGKKIWLGVVTTTSAGGTNN 334  
 Db 495 VDEEAVAQWVSLQTGIPARSARTEASEKLLKLEDLFRKRVIGQNDVAVTSICRAIRRSRTG 554  
 QY 335 LSDSY--AAGFLWLTG-----MLANQ-----GIDVIRHSFFDH----- 368  
 Db 555 IKDPNRPTGSFLFGPTGVGKSLAQAIEMFGGEDALIOVDMSEYMEKFAATRMGSP 614  
 QY 369 -GY-----NHLVQD-----NFNP--LPDYWLSLYK-RL---IGPKVLAV 401  
 Db 615 PGYGVHEGGHLEQVRRRPPCVVLFDEIEKAHPDMDLMQLQILEQGLTDSFGRRKVDPR 674  
 QY 402 HVAGLQRKPRGRVIRDKIRIYAHCTNHHNHNHYVRGSIITLFIINLRSRKKIKLAGTLRD 461  
 Db 675 HAIIMTSLNGLADLRKSGEIGFLKSHMDYKVIQEKIE-----HAMKKHLK--PEFIN 726  
 QY 462 KLVHQLLOPGQEG-----LKSKSVQINGPLV---MVDDGTLPKLPKR 503  
 Db 727 RLDESIVPRPLEKESLSBIIHLEINKLSRLKNYOMALNPDSVTSFLVTKGHSPEMGAR 786  
 QY 504 PLR 506  
 Db 787 PLR 789

RESULT 6  
 RPOC\_CAMJE STANDARD; PRT; 1517 AA.  
 AC Q9PI30;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase  
 beta' chain) (RNA polymerase beta' subunit).  
 GN RPOC OR C30479.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 OC Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCTC 11168;  
 RX MEDLINE=20150912; PubMed=10688204;  
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
 Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,  
 Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
 Quail M.A., Rajandream A.M., Rutherford K.M., van Vliet A.H.M.,  
 Whitehead S., Barrell B.G.;  
 RA "The genome sequence of the food-borne pathogen Campylobacter jejuni  
 reveals hypervariable sequences."  
 RL Nature 403:665-668(2000).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION

CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA}(N).  
 CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE  
 CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1  
 CC BETA' CHAIN.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AL139075; CAB75117.1; -;  
 CC InterPro; IPR000722; RNA\_pol\_A.  
 CC Pfam; PF00623; RNA\_pol\_A; 1.  
 CC Transfaser; DNA-directed RNA polymerase; Transcription;  
 CC KW Complete proteome.  
 CC SQ SEQUENCE 1517 AA; 168821 MW; 6778D27483AB70EC CRC64;

Query Match 3.5%; Score 97.5; DB 1; Length 1517;  
 Best Local Similarity 20.3%; Pred. No. 26;  
 Matches 111; Conservative 87; Mismatches 188; Indels 161; Gaps 31;  
 QY 46 PLPVDRAAGLKEKTEKILLIDVSTKPNVTVNENFLSLQDPSIIHDGWLDFLSKRLVTTIA 105  
 Db 442 PVMNRA-----PTLHKLSIQAFHEPVLVEGK---AQLHPLVCAAFNADFGDQMAVHP 493  
 QY 106 RGLSPAFLRFGKRTDFLOFONLRNPA--KSRGGPDPYLLKNYEDDIRSDVALDK--Q 161  
 Db 494 --LSEAI--AECKVLMSSMILLPASGKSVTVPSQDMVLGIY-----LSLEKAGA 542  
 QY 162 KGC-KIAQHPDYMLQREKAAQMHLLVLLKEQFSNTYSNLIITENPNYTMHGRVNGSQ 220  
 Db 543 KGSHKICTGIDVEMMALESKCLDIHA-----SIQTMVDGRKITTAGRLIVKSI 591  
 QY 221 LGKDYIQLKSLLOPIRIYRSASLYGPNIGRPNKKNVIALLD-----GPMKVAGSTVDVAVT- 274  
 Db 592 L-PDFVPENSNWKNVLEK-----KKDIAALVDYVYKQGLEITASFDRLEN 635  
 QY 275 --WOHCYIDGRVVKVMDFL---KTRLLDLSQIRKIQKVNTYTPGKKIWLGVVTTTS 328  
 Db 636 LGFEYATKAGISISTADIIVPNDKQKAIIDEAKKQVREIQ---NSYNLG-----LITS 684  
 QY 329 AGGTNNLSDSYAAGFLWLTGLMLANQGDIVVIRHSFFDGHYNHLVDQNFNPLPDYWLSL 388  
 Db 685 GERYNKIID-----IKSTNNVLSKEMMKLVEKDK---EGFN-----SI 720  
 QY 389 LYKRLIGPKVLAVHVA-----GLQKRPGRVIRDKLRIYAHCTNHHNHNHYVRG-SITL 441  
 Db 721 YMMADSGARGSAQISQLAAMRGLMTKP-DGSIIITPI-----ISNREGNLVLE 769  
 QY 442 FIINLRSRK-----KIKLAGTLRDKLIVH-----QYLLQPYQEGLSKSVQINGQL 489  
 Db 770 YFISTHGARKGLADTALKTANAGYLTRKLIDVAQNVKITIEDCGTH-----EGVEINE---- 822  
 QY 490 VMVDDGTLPKLPRLRAGRTL---VIPPVT-----MGFFVVKNNV- 527  
 Db 823 ITADSSIIITLEERIL--GRVLAEDVIDPITNSVLVFAEGLTMDDEKAKILGESGKSVNI 880  
 QY 528 --ALACR 532  
 Db 881 RTPITCK 887  
 RESULT 7  
 DP3A\_UREPA STANDARD; PRT; 969 AA.  
 ID DP3A\_UREPA



DR InterPro: IPR004273; Dynein\_heavy.  
DR InterPro: IPR001482; GSPIL\_E.  
DR Pfam: PF03028; Dynein\_heavy\_1.  
DR Pfam: PF00437; GSPIL\_E\_1.  
KW Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.  
FT DOMAIN 676 693 COILED COIL (POTENTIAL).  
FT DOMAIN 1176 1215 COILED COIL (POTENTIAL).  
FT DOMAIN 1327 1351 COILED COIL (POTENTIAL).  
FT DOMAIN 1557 1574 COILED COIL (POTENTIAL).  
FT DOMAIN 1637 1668 COILED COIL (POTENTIAL).  
FT DOMAIN 2045 2073 MICROTUBULE-BINDING (POTENTIAL).  
FT DOMAIN 2195 2218 COILED COIL (POTENTIAL).  
FT DOMAIN 3193 3296 COILED COIL (POTENTIAL).  
FT DOMAIN 3423 3481 COILED COIL (POTENTIAL).  
FT DOMAIN 3778 3809 COILED COIL (POTENTIAL).  
FT NP\_BIND 1943 1950 ATP (POTENTIAL).  
FT NP\_BIND 2240 2247 ATP (POTENTIAL).  
FT NP\_BIND 2605 2612 ATP (POTENTIAL).  
FT NP\_BIND 2947 2954 ATP (POTENTIAL).  
SQ SEQUENCE 4367 AA; 495568 MW; 1E10F3E2D170D6DF CRC64;

Query Match 3.4%; Score 96; DB 1; Length 4367;  
Best Local Similarity 21.1%; Pred. No. 1.5e+02;  
Matches 110; Conservative 57; Mismatches 179; Indels 176; Gaps

QY	33	LHLSLSQSAGDRRLP-----PVDRAAGLKEKTLILDVSTKNPVTVNENFLSLQLDPSSI	87
		:    :    :    :    :    :    :    :    :	
Dd	3548	LHLS-GVFQKHNPWTEYLSTADERSWGENTLPVDDICTEMAILKKFNRYPLIIDPSG	3606
QY	88	IHDGWLDLFSSKRVLARGLSPAF-----LRFGGKRTDLPQNLNRPNKASRGSGPGP	140
		:    :    :    :    :    :    :    :    :	
Dd	3607	RATEEFNRESKDKRLIVTSFLDSDSFVKLESSLRG-----NPILIODAEHL	3653
QY	141	DYYL-----KNYEDDIVRSDVALDGOKGCKIAHQHPDVMLELQ-REKAQMHLVLLKEQFS	194
		:    :    :    :    :    :    :    :    :	
Dd	3654	DPVLNHVLANKEYQTGGRVLIOLGKO---QIDFSPAFLYLSTRDSPSATFPADICSR---	3707
QY	195	NTYSNLITLTPNNRYTMHGNAVNGSQLGDYIQLKSLLQPIRIYSRASLYGNIGRPKN	254
		:    :    :    :    :    :    :    :    :	
Dd	3708	TTFNVFTVTOSS-----LTQSLENEVLKSER-----PDVDERSRN	3742
QY	255	VI-----ALIDGFMKVAGSTVDAVTWQHGYIDGRVVKVMDFLKRLDLTL	299
		:    :    :    :    :    :    :    :    :	
Dd	3743	LIKLOGEFKVHLRQLEKLLQALNESRGNIID-----DDHVLETLTKEAAEIS	3793
QY	300	SDQIRKIQKVNTYTGGKIIWEGVVTTSAGGT---NNLSDSYAAGFWLWTLGLMANQG	356
		:    :    :    :    :    :    :    :    :	
Dd	3794	A-----KMSNT-----EGVNAEVEQITLQYNIIARSCSAVFAVLEQL-----	3830
QY	357	IDVVIRHSFPHGCYNHLVDQNFPDYWLSSLYKRLIGPKVLAVHVAGLORKPPGRVI	416
Dd	3831	-----HYLNHF-----YRFSIQY-----FLDTFHSVLRGNP-----	3856
QY	417	RDKLRIYAHCTNNHNHYRGSIT--LFTINLHRSRKKIKLAGTLRDLKVHGYLL---QP	471
		:    :    :    :    :    :    :    :    :	
Dd	3857	-----HLANETNHNVRDIIIVKDLFYATF----KRTALGLLKQDRITLAMLLAQASP	3904
QY	472	YGQE-GL-----KSKSVQLNGQPVLVMVDGTTPELK	501
		:    :    :    :    :    :    :    :    :	
Dd	3905	YKMDKGLLDIILDERIEGKDVSIDQNTREEAFARAKKIPALK	3946

  

RESULT	10				
SNK_RAT	ID	SNK_RAT	STANDARD;	PRT;	682 AA.
AC	Q9R012;				
DT	16-OCT-2001	(Rel. 40, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Serine/threonine-protein kinase SNK	(EC 2.7.1.-) (Serum inducible kinase).			
DN	KINASE				
DN	SNK				

OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_TaxID=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RP	MEDLINE=9452760; Pubmed=10523297;	
RA	Kauselmann G., Weiler M., Wulff P., Jessberger S., Konietzko U.,	
RA	Scafield J., Staubli U., Bereiter-Hahn J., Streibhardt K., Kuhl D.,	
RT	"The polo-like protein kinases Fnk and Snk associate with a Ca(2+)-and	
RT	integrin-binding protein and are regulated dynamically with synaptic	
RT	plasticity."	
RL	EMBO J. 18:5528-5539(1999).	
CC	-!- FUNCTION: MAY PLAY A ROLE IN THE DIVISION OF AT LEAST SOME CELL	
CC	TYPES, SUCH AS FIBROBLASTS, AND COULD FUNCTION IN EMBRYOGENESIS,	
CC	WOUND HEALING OR NEOPLASIA (BY SIMILARITY).	
CC	-!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.	
CC	CDC5/POLO SUBFAMILY.	
CC	-!- SIMILARITY: CONTAINS 2 POLO BOX DOMAINS.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL; AF136583; AAF08366.1; -.	
DR	HSSP; P00518; 1PHK.	
DR	InterPro: IPR000719; Euk_pkinase.	
DR	InterPro: IPR000959; POLO_box.	
DR	InterPro: IPR002230; Ser_thr_pkinase.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF00659; POLO_box; 2.	
DR	SMART; SM00220; S_TKC; 1.	
DR	PROSITE; PS50078; POLO_BOX; 2.	
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.	
DR	PROSITE; PS500108; PROTEIN_KINASE_ST; 1.	
KW	Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat.	
FT	DOMAIN 54 59 POLY-HIS.	
FT	DOMAIN 79 331 PROTEIN KINASE.	
FT	NP_BIND 85 93 ATP (BY SIMILARITY).	
FT	BINDING 108 108 ATP (BY SIMILARITY).	
FT	ACT_SITE 202 202 BY SIMILARITY.	
FT	DOMAIN 507 570 POLO_BOX 1.	
FT	DOMAIN 603 674 POLO_BOX 2.	
SQ	SEQUENCE 682 AA; 77919 MW; 58C50DEBDE83D5F3 CRC64;	
-----		
Query Match 3.4%; Score 94.5; DB 1; Length 682;		
Best Local Similarity 22.6%; Pred. No. 15;		
Matches 95; Conservative 58; Mismatches 166; Indels 101; Gaps 21;		
Qy	129 RNPASRGPGPDYLYKNYEDDIVRSVDALDKQCKIAQHPDVMLELQREKAAQ-----183	
Db	81 RGVVLGKG-----FACYE-----WDLTNKKVYAALIPHSRVAPHQREKIDKEILH 131	
Qy	184 -----MHLVLKLEQFSNTYSLNLTPENNYRTMHGRAVNGSQLGKDYIQLKSLLOP-IRI 237	
Db	132 RILHKKHVQFHYFEDKENIYILLEYCSRSM-----AHLK---ARKVLTEPEVRY 181	
Qy	238 YSRASLQPGNIGRPNKKNVIALLDGFMKAVAGSTVDVATWQHCHYIDGRVVKVMDFLKLRLD 297	
Db	182 YLRQIVSGLKYLHGEQ-----ILHRDLKLGNNFFIN-----EAMELKVGDFGLAARLE 228	
Qy	298 TLDQIRKIKQVWNTYTP-----GKKIWLEGVV---TTSAG-----GTNNLSDSY--- 339	
Db	229 PLEHRRRTICGTPVLSPEVLNKGHCESDIWALGCMVTMLLIGRPPFTTNLKETYRC 288	
Qy	340 --AGFLWLNTL-----GMLANQ-----GIDVVIRHSFEDHGYY--NHLVDQNFNPL 381	

Db	289 IREARYTMPSSLAPAKHLIASMLSKNPEDRPSLDDIIRHDFLQGTTPDRLSSSCCHTV 348	
Qy	382 PDYWLSLLYRLGPKVLAVHVGLOKQPR---PGRVIRDKLRIYAHCTNHNHNHYVRG 437	
Db	349 PDFHLSSPAKNFF-KKAAALFGGKKDKARYNDTHNKVSKEDDIY-----KLRLDLKKT 402	
Qy	438 SITLFIINLHRSRKKIKLACTLRDLKLVHVOYLLQPYGOGLKSKSVQINGOPLVWVDGTL 497	
Db	403 SITQ-QFSKRTDEELQPPPTFAK-----SGTSAVENK--QQIGDAIRMIVRGTL 450	
RESULT 11		
SNK_HUMAN		
ID	SNK_HUMAN	STANDARD; PRT; 685 AA.
AC	Q9NYX3; O60679; Q9UE61;	
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Serine/threonine-protein kinase SNK (EC 2.7.1.-) (Serum inducible kinase).	
GN	SNK.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Ouyang B., Dai W.,	
RL	Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.	
RL	[2]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Colon;	
RA	Anderson K.M., Nerurkar S.S., Hansbury M.J., Fornwald J., Scott G.,	
RA	Bouzyk M., Mui P., Imbruglia C.S., Carlson K., Marshall L.A.,	
RA	Roshak A.K.	
RT	"Identification and characterization of human serum-inducible kinase	
RT	(SNK), a novel member of the polo-kinase family of cell cycle	
RT	regulators: potential implication for regulation of vascular smooth	
RT	muscle proliferation."	
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.	
RL	[3]	
RP	SEQUENCE OF 110-408 FROM N.A.	
RA	Fidler C., Boulwood J., Wang Jabs E., Wainscoat J.S.;	
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.	
CC	-!- FUNCTION: MAY PLAY A ROLE IN THE DIVISION OF AT LEAST SOME CELL	
CC	TYPES, SUCH AS FIBROBLASTS, AND COULD FUNCTION IN EMBRYOGENESIS,	
CC	WOUND HEALING OR NEOPLASIA (BY SIMILARITY).	
CC	-!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.	
CC	CDC5/POLO SUBFAMILY.	
CC	-!- SIMILARITY: CONTAINS 2 POLO BOX DOMAINS.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL; AF059617; AAC14573.1; -.	
DR	EMBL; AF223574; AAF62897.1; -.	
DR	EMBL; U85755; AAD00375.1; -.	
DR	HSSP; P00518; 1PHK.	
DR	InterPro: IPR000719; Euk_pkinase.	
DR	InterPro: IPR000959; POLO_box.	
DR	InterPro: IPR002230; Ser_thr_pkinase.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF00659; POLO_box; 2.	
DR	SMART; SM00220; S_TKC; 1.	
DR	PROSITE; PS50078; POLO_BOX; 2.	
DR	PROSITE; PS500107; PROTEIN_KINASE_ATP; 1.	
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.	
DR	PROSITE; PS500108; PROTEIN_KINASE_ST; 1.	

```
KW Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat.
FT DOMAIN 57 64 POLY-HIS.
FT NP_BIND 82 334 PROTEIN KINASE.
FT BINDING 88 96 ATP (BY SIMILARITY).
FT ACT_SITE 111 111 ATP (BY SIMILARITY).
FT DOMAIN 205 205 BY SIMILARITY.
FT NP_BIND 510 573 POLO BOX 1.
FT DOMAIN 510 573 POLO BOX 2.
FT CONFLICT 28 28 G -> A (IN REF. 2).
SQ SEQUENCE 685 AA; 78222 MW; DECBE929CA35A412 CRC64;

Query Match 3.4%; Score 94.5; DB 1; Length 685;
Best Local Similarity 23.1%; Pred. No. 15;
Matches 84; Conservative 46; Mismatches 144; Indels 89; Gaps 18;

QY 129 RNPASRGPGPDYLLKNEYDIVRSVALDKQCKIAQHPDVMLQREKAAQ-----183
DB 84 RGVLGKGG-----FAKCYE-----WTDLTNNKYVAAKLIIPHSRVAKPHQREKIDKEIELH 134
QY 184 -----MHVLVLKEQFSNTYSNLIITPEPNRYTMHGRAVNGSQLGKDYIQLKSLQIP-IRI 237
DB 135 RILHHKVVQVHYVEDKENIYVILLEYCSRSM-----AHILK-----ARKVLTEPEVRY 184
QY 238 YSRASLYGPNIGRPRKKNVIALLDGPMKVAGSTVDVATVQHCYIDGRVVKVMDFLKTRLLD 297
DB 185 YLRQIVSGIKYLHEQE-----ILHRDLKLGNNFIN-----EAMELKVGDFGLAARLE 231
QY 298 TLDQIRKIQKVVNTYTP-----GKKIWLGVV--TTSAG-----GYNLSDSY-- 339
DB 232 PLEHRRVTCGPNVLSPEVLNKGQHGCESDIWLGCWYMTLGRPPFTTNLKETRC 291
QY 340 --AAGFLWNL-----GMLANQ-----GIDVIRHSFFDHGY--NHLVDQNFNPL 381
DB 292 IREARYTMPSSILLAPAKHLLIASMLSKNPEDRPSLDIIRHDFLQGTPTDRLSSCCHTV 351
QY 382 PDYWLSLLYKRLGPKVLAVHAGLQRPKPR---PGRVIRDKLRIYACTNHHNNHYRG 437
DB 352 PDFHLVSPAKNFF-KKAAALFGGKKDKARYIDTHNRVSKEDEDIY-----KLRHDLKKT 405
QY 438 SIT 440
DB 406 SIT 408

RESULT 12
HELX_METTH STANDARD; PRT; 690 AA.
AC 026901;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Putative ski2-type helicase (EC 3.6.1.-).
GN MTH810.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OX Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -I- SIMILARITY: BELONGS TO THE HELICASE FAMILY. SKI2 SUBFAMILY.
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000858; AAB85310.1; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR003583; HHH1.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD; 1.
DR SMART; SM00271; Helicase_C; 1.
DR SMART; SM00490; HELICC; 1.
DR SMART; SM00278; HHH1; 1.
KW Hypothetical protein; Hydrolase; Helicase; ATP-binding;
FT NP_BIND 45 52 ATP (POTENTIAL).
FT SITE 133 136 DEXH BOX.
SQ SEQUENCE 690 AA; 78509 MW; 90BEE6110C4133F3 CRC64;

Query Match 3.4%; Score 94.5; DB 1; Length 690;
Best Local Similarity 21.8%; Pred. No. 15;
Matches 79; Conservative 52; Mismatches 112; Indels 119; Gaps 18;

QY 22 LAPGALYLALLHLHLS-LSSQAG-----DRRLPLVDRAA-----GLKEKTLILLDY 65
DB 157 LNPVSRIVALSATLSNMDIAGWLDARVVEHDYRPVPLHREVLDTMFQVREKNDVWLK 216
QY 66 STKPNVTVNENFLSLQDPSIIHDG--WLDFLSKRLV-LTARGLSPAFLRFGKRTDF 122
DB 217 LERS-----LEDGSQLAFVSTRTTESLASHAD---KISKIPDD 255
QY 123 LOFQNLNRP-----KSRGGPGDYLLKNEYDIVRSVALDKQCKIAQHPDVML 175
DB 256 M-VESFREYVAGVLEVPKSRGSPPTSTCLK-----LAECLEAGIAFHAGLFN 302
QY 176 LOREKAAQMHVLVLKEQFSNTYSNLIITPEPNRYTMHGRAVNGSQLGKDYIQLKSLQPI 235
DB 303 QRRE-----IIEDEFQGNILMITATPS---LMYG-----VNLPSRTVVI 339
QY 236 RIYSRASLYGP-----NIGRPRKNVIALLDGF-MKVAGSTVDVATVQHCYID 281
DB 340 RDTWTWTSQPRIPVFDYEQMSGRAGRPYDDA---GYSYLIARSHDEAMDLEYYIR 395
QY 282 GRVVKVMDFLKTRLLDITLSDQIRKIQKVVNTYTPGKKIWLGVVTSAGTNNLSDSYAA 341
DB 396 GEVERT---TSRIENRDALYRQI-----IAQVAGSLSGTTEELADFFRN 437
QY 342 GF 343
DB 438 TF 439

RESULT 13
POL_SIVMK STANDARD; PRT; 1054 AA.
AC P05697;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE POL polyprotein (Contains: Protease (Retropepsin) (EC 3.4.23.-);
DE Reverse transcriptase (EC 2.7.7.49); Ribonuclease H (EC 3.1.26.4)).
GN POL.
OS Simian immunodeficiency virus (K6W isolate) (SIV-MAC).
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11735;
RN [1]
RP SEQUENCE FROM N.A.
```



```
Qy 414 --RVIRDKLRIYAHCTNNHNNYVRGSTITFLINLHRS---RKKIKLAG-----TLRDKL 463
Db 235 DENYLKDIQPCWLTVTYTTENSITLRG-----NLHRAPLFGIVKGVPRYPCPSIEDKI 287
Qy 464 V-----HOYLLQPYGOEGLKSKSVQLNGQPLVMVDGCTLPKLPRLRAGRITVIPPV 516
Db 288 TRFADKPRHQFLFLEP---EGRNTEEVYIGGLSTSPMEDVQFDLVKSIPLGLENAQMMPGY 344
Qy 517 TMGFEWV 523
Db 345 ALEYDVV 351

RESULT 15
SYV2_BACSU
ID SYV2_BACSU STANDARD; PRT; 412 AA.
AC P25151;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tyrosyl-tRNA synthetase 2 (EC 6.1.1.1) (Tyrosine--tRNA ligase)
DE (TYRS 2).
CN TYR2 OR TYR OR TYRS1 OR TYRT OR IPA-9R.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=92216127; PubMed=1806041;
RA Glaser P., Kunst F., Debarbouille M., Vertes A., Danchin A.,
RA Dedonder R.;
RT "A gene encoding a tyrosine tRNA synthetase is located near sacs in
RT Bacillus subtilis".
RL DNA Seq. 1:251-261(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Prescan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
CC -!- CATALYTIC ACTIVITY: ATP + L-tyrosine + tRNA(Tyr) = AMP +
CC diphosphate + L-tyrosyl-tRNA(Tyr).
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52480; CAA36724.1;
DR EMBL; X73124; CAA51565.1; ALT_INIT.
DR EMBL; Z99123; CAB15872.1; ALT_INIT.
DR PIR; S16426; S16426.
DR PIR; S39664; S39664.
DR HSP; P00952; 4TS1.
DR Subtilist; BG10555; tyr2.
DR InterPro; IPR002942; S4.
DR InterPro; IPR002305; tRNA-synt_lb.
DR InterPro; IPR001412; tRNA-synt_I.
DR InterPro; IPR002307; tRNA-synt_tyr.
```

```
DR Pfam; PF01479; S4; 1.
DR Pfam; PF00579; tRNA-synt_lb; 1.
DR PRINTS; PR01040; TRNASYNTHYR.
DR SMART; SM00363; S4; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Repeat; Complete proteome.
FT SITE 57 66 "HIGH" REGION.
FT SITE 241 245 "KMSKS" REGION.
FT BINDING 244 244 ATP (BY SIMILARITY).
FT DOMAIN 88 100 2 X 6 AA TANDEM REPEATS.
FT REPEAT 88 93 1.
FT REPEAT 95 100 2.
SQ SEQUENCE 412 AA; 46969 MW; DDF24A1A6E683441 CRC64;

Query Match 3.4%; Score 93.5; DB 1; Length 412;
Best Local Similarity 21.1%; Pred.No.8.6; Indels 87; Gaps 21;
Matches 70; Conservative 67; Mismatches 107;

Qy 28 YLALLHLSSLQAGDRRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNEFLSLQLDPSI 87
Db 81 HIVQLLIGDFTGKIGD---PTGKSAARKQ-----LTDEEVOHNAKTYPEQFGKV-LDPEK 131
Qy 88 I---HDGWLDFLSSKRLVTLARGLSPAFLRFGGKRTDFLQFQNLNRPASRGSGPPDY 143
Db 132 VELHYNKWLKTLNLEDVIELAGKITVARLM---ERDDFEERIAMQKPISLH-----EFF 183
Qy 144 ---LKNYEDDIVRSDVALDKQKGCKIAQHPDVM-----ELQREKAAQMHVLVK---- 190
Db 184 YPLMQGYDSVWLESDELGGTD-----QHNVLGRHFQERYNKEQKVVLMPLEGLDG 238
Qy 191 -RQFSNTYSNLI-LTE-PNNYRTMHGRAVN-GSQLCKDYIOLKSL-----OPTRIYSR 240
Db 239 VEKMSKSKHNYIGINEHPND---MYGKTNLSPDSLMMKKYIHLATDLELEKKQLVKDLET 295
Qy 241 ASLYGPNIGRPR-----KNVIALDGMKPVAGST-----VDVATWQH 277
Db 295 GAVH-----PRDAKMLLARTIVRMVHG-EKAAEAAEHFSKTVFQENSLPEDIPAVNMK- 347
Qy 278 CYIDGRVVKVMDFL-KTRLDDTLSDQIRKIQ 307
Db 348 ---GEKTAMIDLLVKLLSLSSKSEARRMIQ 375

Search completed: July 30, 2002, 08:31:25
Job time: 972 sec
```

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:14:08 ; Search time 123.94 Seconds  
(without alignments)  
745.355 Million cell updates/sec

Title: US-09-836-461-2  
Perfect score: 2785  
Sequence: 1 MRVLCAPFEMPSSNSRPPA.....PVTMGFFVKNVNALACRYR 534

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2785	100.0	534	4	Q9HB38 homo sapien
2	2746	98.6	592	4	Q9HB37 homo sapien
3	2478	89.0	480	4	Q9HB39 homo sapien
4	953	34.2	536	11	Q9QZF8 rattus norv
5	937	33.6	545	4	Q9UL39 homo sapien
6	932.5	33.5	543	4	Q9Y251 homo sapien
7	932	33.5	545	6	Q9MYV0 bos taurus
8	836.5	30.0	523	13	Q90YK5 gallus gall
9	302	10.8	521	10	Q9SDA1 arabidopsis
10	302	10.8	543	10	Q9FFI0 arabidopsis
11	261.5	9.4	516	10	Q9FLK8 arabidopsis
12	256.5	9.2	527	10	Q9LRC8 arabidopsis
13	246	8.8	536	10	Q9FZP1 arabidopsis
14	171	6.1	935	5	Q9VE79 drosophila
15	155	5.6	190	10	O82604 arabidopsis
16	113	4.1	4533	5	Q9BIX3 tetrahymena

ALIGNMENTS

RESULT 1

ID	Q9HB38	PRELIMINARY;	PRT;	534 AA.
AC	Q9HB38;			
DT	01-MAR-2001 (TREMREL. 16, Created)			
DT	01-MAR-2001 (TREMREL. 16, Last sequence update)			
DT	01-MAR-2001 (TREMREL. 16, Last annotation update)			
DE	HEPARANASE-LIKE PROTEIN HPA2B.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20483645; PubMed=11027606;			
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,			
RA	Hircock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;			
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian			
RT	Heparanase Family Member."			
RL	Biochem. Biophys. Res. Commun. 276:1170-1177(2000).			
DR	EMBL; AF282886; AAG23422.1; -			
SQ	SEQUENCE 534 AA; 60063 MW; C3DE5E900CB338C4 CRC64;			

Query Match	100.0%;	Score 2785;	DB 4;	Length 534;
Best Local Similarity	100.0%;	Pred. No. 8e-216;		
Matches 534;	Conservative	0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	1	MRVLCAPFEMPSSNSRPPACIALYALLHLHLSLSQAGDRPLPVDRAAGLKEKTL 60		
Db	1	MRVLCAPFEMPSSNSRPPACIALYALLHLHLSLSQAGDRPLPVDRAAGLKEKTL 60		
Qy	61	ILLDVSTKNPVRTVNFENLSLQDPSIIHDGWLDFSSKRLVTLARGLSAPFLREGGRT 120		
Db	61	ILLDVSTKNPVRTVNFENLSLQDPSIIHDGWLDFSSKRLVTLARGLSAPFLREGGRT 120		
Qy	121	DFLQFQNLNPAKSGGPGPDYLYKNYEDDIVRSVALDKQKCKIAQHPDVMLEQREK 180		
Db	121	DFLQFQNLNPAKSGGPGPDYLYKNYEDDIVRSVALDKQKCKIAQHPDVMLEQREK 180		

```
QY 181 AAQHLVLLKEQFSNTYSNLILTEPNNTYTMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 240
DB 181 AAQHLVLLKEQFSNTYSNLILTEPNNTYTMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 240
QY 241 ASLYGPNIGRPRKNNVIALLDGFMKVAGSTVDVATWQHICYIDGRVVKVMDFLKTRLLDLS 300
DB 241 ASLYGPNIGRPRKNNVIALLDGFMKVAGSTVDVATWQHICYIDGRVVKVMDFLKTRLLDLS 300
QY 301 DOIRKIQVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWNLTLGLMANGIDVV 360
DB 301 DOIRKIQVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWNLTLGLMANGIDVV 360
QY 361 IRHSFDFHGYNHLVDQNFNPLDYWLSLLYKRLIGPKVLAVHVGAGLQRPGRVIRDKL 420
DB 361 IRHSFDFHGYNHLVDQNFNPLDYWLSLLYKRLIGPKVLAVHVGAGLQRPGRVIRDKL 420
QY 421 RIYACTNHHNNHNVYRGSIITLFIINLHRSRKKIKLAGTLRDKLVHQLVYLLQPYGQGLKSK 480
DB 421 RIYACTNHHNNHNVYRGSIITLFIINLHRSRKKIKLAGTLRDKLVHQLVYLLQPYGQGLKSK 480
QY 481 SVQLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFFVVKVNNALACRYR 534
DB 481 SVQLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFFVVKVNNALACRYR 534

RESULT 2
QY 09HB37 PRELIMINARY; PRT; 592 AA.
AC 09HB37;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE HEPARANASE-LIKE PROTEIN HPA2C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20483645; PubMed=11027606;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian
RT Heparanase Family Member.";
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).
DR EMBL; AF282887; AAG33423.1; -.
SQ SEQUENCE 592 AA; 66580 MW; 95C384AD9A74258E CRC64;
```

```
Query Match 98.6%; Score 2746; DB 4; Length 592;
Best Local Similarity 90.2%; Pred. No. 1.3e-212;
Matches 534; Conservative 0; Mismatches 0; Indels 58; Gaps 1;

QY 1 MRVLCAPPEAMPSSNRPPACIAPGALYLAALLHLSSSQAGDRRLPVDRAAGLKEKTL 60
DB 1 MRVLCAPPEAMPSSNRPPACIAPGALYLAALLHLSSSQAGDRRLPVDRAAGLKEKTL 60
QY 61 ILDDYSTKNPVRTVNEFLSLQDPSIHDGWLDFLSKRLVTLARGLSPAFLRFGGKRT 120
DB 61 ILDDYSTKNPVRTVNEFLSLQDPSIHDGWLDFLSKRLVTLARGLSPAFLRFGGKRT 120
QY 121 DFLOFQNLRLNPAKSRGGPGPDYILKNYEDDIVRSDVALDKQKCKIAQHPDVMLELQREK 180
DB 121 DFLOFQNLRLNPAKSRGGPGPDYILKNYEDDIVRSDVALDKQKCKIAQHPDVMLELQREK 180
QY 181 AAQHLVLLKEQFSNTYSNLILTEPNNTYTMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 240
DB 181 AAQHLVLLKEQFSNTYSNLILTEPNNTYTMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 240
QY 204 -----EPNNYRTMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 242
DB 241 ALSLLKYSASKKYNISWELGNENPNYRTMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 300
```

```
QY 243 LYGPNIGRPRKNNVIALLDGFMKVAGSTVDVATWQHICYIDGRVVKVMDFLKTRLLDLSQ 302
DB 301 LYGPNIGRPRKNNVIALLDGFMKVAGSTVDVATWQHICYIDGRVVKVMDFLKTRLLDLSQ 302
QY 303 IRKQKVVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWNLTLGLMANGIDVVIR 362
DB 361 IRKQKVVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWNLTLGLMANGIDVVIR 362
QY 363 HSFDHGYNHLVDQNFNPLDYWLSLLYKRLIGPKVLAVHVGAGLQRPGRVIRDKLRI 422
DB 421 HSFDHGYNHLVDQNFNPLDYWLSLLYKRLIGPKVLAVHVGAGLQRPGRVIRDKLRI 422
QY 423 YAHCTNHHNNHNVYRGSIITLFIINLHRSRKKIKLAGTLRDKLVHQLVYLLQPYGQGLKSK 482
DB 481 YAHCTNHHNNHNVYRGSIITLFIINLHRSRKKIKLAGTLRDKLVHQLVYLLQPYGQGLKSK 482
QY 483 QLNQGPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFFVVKVNNALACRYR 534
DB 541 QLNQGPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFFVVKVNNALACRYR 592

RESULT 3
QY 09HB39 PRELIMINARY; PRT; 480 AA.
AC 09HB39;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE HEPARANASE-LIKE PROTEIN HPA2A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20483645; PubMed=11027606;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian
RT Heparanase Family Member.";
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).
DR EMBL; AF282885; AAG23421.1; -.
SQ SEQUENCE 480 AA; 53900 MW; F75F89F67AC1FF83 CRC64;
```

```
Query Match 89.0%; Score 2478; DB 4; Length 480;
Best Local Similarity 89.9%; Pred. No. 3.6e-191;
Matches 480; Conservative 0; Mismatches 0; Indels 54; Gaps 1;
```

```
QY 1 MRVLCAPPEAMPSSNRPPACIAPGALYLAALLHLSSSQAGDRRLPVDRAAGLKEKTL 60
DB 1 MRVLCAPPEAMPSSNRPPACIAPGALYLAALLHLSSSQAGDRRLPVDRAAGLKEKTL 60
QY 61 ILDDYSTKNPVRTVNEFLSLQDPSIHDGWLDFLSKRLVTLARGLSPAFLRFGGKRT 120
DB 61 ILDDYSTKNPVRTVNEFLSLQDPSIHDGWLDFLSKRLVTLARGLSPAFLRFGGKRT 120
QY 121 DFLOFQNLRLNPAKSRGGPGPDYILKNYEDDIVRSDVALDKQKCKIAQHPDVMLELQREK 180
DB 121 DFLOFQNLRLNPAKSRGGPGPDYILKNYEDDIVRSDVALDKQKCKIAQHPDVMLELQREK 180
QY 181 AAQHLVLLKEQFSNTYSNLILTEPNNTYTMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 240
DB 150 -----EPNNYRTMHGRAVNSQLGKDYIQLKSLLOPIRIYSR 186
QY 241 ASLYGPNIGRPRKNNVIALLDGFMKVAGSTVDVATWQHICYIDGRVVKVMDFLKTRLLDLS 300
DB 187 ASLYGPNIGRPRKNNVIALLDGFMKVAGSTVDVATWQHICYIDGRVVKVMDFLKTRLLDLS 246
QY 301 DOIRKIQVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWNLTLGLMANGIDVV 360
DB 247 DOIRKIQVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWNLTLGLMANGIDVV 306
```

QY 361 IRHSFFDHGYNHLVDQNFPLPDYWLSSLYKRLIGPKVLAVHVGAGLQRPGRVIRDKL 420  
|||||  
Db 307 IRHSFFDHGYNHLVDQNFPLPDYWLSSLYKRLIGPKVLAVHVGAGLQRPGRVIRDKL 366  
|||||  
QY 421 RIYAHCTNNHHNNYVRGSIITLFIINLHRSRKKTKLAGTLRDKLVHQLYLLQPYGOEGLSK 480  
|||||  
Db 367 RIYAHCTNNHHNNYVRGSIITLFIINLHRSRKKTKLAGTLRDKLVHQLYLLQPYGOEGLSK 426  
|||||  
QY 481 SVOLNGOPLVMVDGTLPELKPRLRAGRTLVIPTVTMGFFVVKVNNALACRYR 534  
|||||  
Db 427 SVOLNGOPLVMVDGTLPELKPRLRAGRTLVIPTVTMGFFVVKVNNALACRYR 480  
|||||  
RESULT 4  
Q9QZF8 PRELIMINARY; PRT; 536 AA.  
AC Q9QZF8;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE HEPARANASE.  
GN HEP.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Podyma K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;  
RT "Heparanase from parathyroid cell line."  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF184967; AAF04563.1; -  
SQ SEQUENCE 536 AA; 60569 MW; 6208B1FD9EE28421 CRC64;

Query Match 34.2%; Score 953; DB 11; Length 536;  
Best Local Similarity 39.3%; Pred. No. 3e-68;  
Matches 216; Conservative 83; Mismatches 165; Indels 86; Gaps 10;

QY 42 GDRRLPVDRAAG-LKEKTLILLDVSTKNPVRTVNENFSLQLDPSIIHD-GWLDLSSK 99  
|||  
Db 12 GLRLALTOGTAGTATKDVVDLEFFYKRLFQSVSPSFLSIITDASLATDPRFLTLGSP 71  
|||  
QY 100 RLVTLAGLSAPLRGGKRTDFLQFONLRNPAKSGGPGDYLYKNYEDDIVRSVDALD 159  
|||  
Db 72 RLRLALAGLSAPLRGGKRTDFLFI-----DPNKEPTSEERSYMQSDNDICGSE--- 123  
|||  
QY 160 KQGGCKIAQHPVMLELQREKAAQMHVLLKEQF-----SNTYS----- 198  
|||  
Db 124 -----RVSADVLRLQMEWPFQ-ELLRLREQYQREFKNSTYSRSSVDMLYSFACKSRL 175  
|||  
QY 199 -----NLIL-----TEPNYRTMHGRAVNGSQL 221  
|||  
Db 176 DLIFGLNALLRTPDLRWNSSNAQLLNYCSSKGYNTISWELGNPNFQWKAQISIDGLQL 235  
|||  
QY 222 GKDYIQLKSLLOPIRIYSRASLYGPNIGRPKNVIALIDGFMKAVGSTVDVATWQHCYID 281  
|||  
Db 236 GEDFVELHKLQK-SAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGBEVIDSLTWHHYLN 294  
|||  
QY 282 GRVVKVMDFLKTLRLDLSQIRKIQKVNTYTPGKKIWLEGVVTYSAGGTNNLSDSYAA 341  
|||  
Db 295 GRVATKEDFLSSVDLDTFLISVQIKLVTKEMTPGKKVWLGSETSSAYGGGAPLLSNTFAA 354  
|||  
QY 342 GFLWNLATGLMANOGIDVIRHSFFDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAV 401  
|||  
Db 355 GFNWLKGLSAGLGLGVNMRQVFFGAGNYHLVDENFEPLDYLWLSLLFKLVGPKVLM 414  
|||  
QY 402 HVAGLQRPGRVIRDKLRIYAKRHHNNHNNYVRGSIITLFIINLHRSRKKIKLAGTLRD 461  
|||  
Db 415 RVKGPD-----RSKRLVLRCHTNVYHPRYREGDLTVLVNLHNVTKHLKLPDPMFS 465  
|||  
QY 462 KLVHQLYLLQPYGOEGLSKSVOLNGOPLVMVDGTLPELKPRLRAGRTLVIPTVTMGFF 521  
|||

Db 466 RPVDKYLKPFSGDGLLSKSVOLNGOPLVMVDGTLPELTKPLPAGSSLSVAFSYGFF 525  
|||||  
QY 522 VKNNVNALAC 531  
|||  
Db 526 VIRNAKIAAC 535  
|||  
RESULT 5  
Q9UL39 PRELIMINARY; PRT; 545 AA.  
AC Q9UL39;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE HEPARANASE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RX MEDLINE=20229546; PubMed=10764835;  
RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;  
RT "Heparanase expression in invasive trophoblasts and acute vascular  
RT damage."  
RL Glycobiology 10:467-475(2000).  
DR EMBL; AF084467; AAD54516.1; -  
SQ SEQUENCE 545 AA; 61418 MW; 67B80ACD73C5A9A1 CRC64;

Query Match 33.6%; Score 937; DB 4; Length 545;  
Best Local Similarity 38.1%; Pred. No. 5.9e-67;  
Matches 219; Conservative 78; Mismatches 184; Indels 94; Gaps 11;

QY 17 RPPACIAPGALYIALLLHLSLSSOAGDRRLPVDVRAAGLKEKTLILLDVSTKNPVRTVNE 76  
|||  
Db 4 RSPALPPLPPLMLLLGLPLGPLSPGAPRAQA-----QDDVVDLDFTQEPHLVSP 56  
|||  
QY 77 NELSLOLDPSIIHD-GWLDLSSKRLVTLAGLSAPLRFEGKRTDFLQFONLRNPAKSR 135  
|||  
Db 57 SFLSVITIDANLADTPREFIILGSPKRLTARGUSPAYLRFGGTKDFLI-----DPKES 112  
|||  
QY 136 GGPDPDYLYKNYEDDIVRSVDALDKQKCKIAQ-HPDVMLELQREKAAQMHVLLKQF- 193  
|||  
Db 113 TFEERSYMQSQVQNDI-----CKYGSIPDPVEEKLRLWEYQEQOL-LLREHYQ 159  
|||  
QY 194 -----SNTYS-----NLIL----- 202  
|||  
Db 160 KFKKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTPADLQWNSNAQLLDLYCSSKGYN 219  
|||  
QY 203 -----TEPNYRTMHGRAVNGSQLGKDYIQLKSLLOPIRIYSRASLYGPNIGRPKNVI 256  
|||  
Db 220 ISWELGNPNFSLKADIFINGSQLGDFIQLHKLRLK-STFKNAKLYGPDVGPQRRTA 278  
|||  
QY 257 ALLDGFMKVAGSTVDVATWQHCYIDGRVVKVMDFLKTLRLDLSQIRKIQKVNTYTPG 316  
|||  
Db 279 KMLKSLKAGGEVIDSVTWHHYLNGRTATREDFLNPDLVDLDFISSVQKVFQVVESTRPG 338  
|||  
QY 317 KKIWLEGVVTYSAGGTNNLSDSYAAAGFLMNTLGMANQGDIVVIRHSFFDHGYNHLVDQ 376  
|||  
Db 339 KKVWLGSETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGTEVVMQVFFGAGNYHLVDE 398  
|||  
QY 377 NEPLPDYWLSSLYKRLIGPKVLAVHVGAGLQRPGRVIRDKLRIYAHCTNNHHNNHNVYR 436  
|||  
Db 399 NEPLPDYWLSSLYKRLIGPKVLAVHVGAGLQRPGRVIRDKLRIYAHCTNNHHNNHNVYR 449  
|||  
QY 437 GSITLFIINLHRSRKKIKLAGTLRDKLVHQLYLLQPYGOEGLSKSVOLNGOPLVMVDGTT 496  
|||  
Db 450 GDULTYAINLHNNYKYLRLPYPSFNKQVDKYLRLPLGPHGLLSKSVOLNGLTLKQVDDQT 509  
|||  
QY 497 LPELKPRLRAGRTLVIPTVTMGFFVVKVNNALAC 531  
|||



```
Db 218 YNISWELGNEPNSFORAKGIFINGRQLGEDFIEFRKLLGK-SAFKNAKLYGPDIGQPRRN 276
Qy 255 VIALIDGPMKVAGSTVDAVTHQHCYIDGRVVKVMDFLKRLDLSDOIIRKIQKVNTYT 314
Db 277 TVMKLSKFLKAGGEVDSVTHHHYVNGRIATKEDFLNPDILDFISSVQKTLRIVEKIR 336
Qy 315 PGKIKWLEGVVTTSAGGTNNISDSYAAGFLWNTLGLMANGIDVIRHSFFDHGYNHLV 374
Db 337 PLKKVWLGETSFAFGGAPFLSNTFAAGFMWLDKLGLSARMGIEVVMRQVLFAGNYHLV 396
Qy 375 DQNFENPLDYWLSLTKRLGKVLAVHAGLQRPGRVIRKRLIYAHCTNNHNNY 434
Db 397 DQNEPLEDYWLSLTKRLGKVLAVHAGLQRPGRVIRKRLIYAHCTNNHNNY 447
Qy 435 VRGSITLFIINLHRSRKKIKLAGTLRDLKLVHQLQPYQOGLKSKSVQLNGQPLVMVDD 494
Db 448 KEGDLTYALNLHNVTKHLEPHLFENKQVDKYLKPSGTDLGSKSVQLNGQILKMWDE 507
Qy 495 GTLPDLKPLRPRLAGRTLVLPVPTMGFFVVKVNNALAC 531
Db 508 QTLPALTEKPLHPGSSGLGMPFFSYGFFVIRNAKVAAC 544

RESULT 8
Q90YK5 PRELIMINARY; PRT; 523 AA.
AC Q90YK5
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE HEPARANASE.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11387326;
RA Goldsmith O., Zcharia E., Alingorn H., Guatta-Rangini Z., Atzmon R.,
RA Michal I., Pecker I., Mitrani E., Vlodaysky I.;
RT "Expression Pattern and Secretion of Human and Chicken Heparanase Are
RT Determined by Their Signal Peptide Sequence.";
RL J. Biol. Chem. 276:29178-29187(2001).
DR EMBL; AY037007; AAK82648.1; -
SQ SEQUENCE 523 AA; 58386 MW; 8EB087B18C9BF881 CRC64;

Query Match 30.0%; Score 836.5; DB 13; Length 523;
Best Local Similarity 35.7%; Pred. No. 6.9e-59;
Matches 200; Conservative 79; Mismatches 187; Indels 95; Gaps 10;

Qy 27 LYLALLHLSLSSQAGDRRPLPVDRAAGLKEKTLILLDVTKNPVRTVNEFLSLQLDPS 86
Db 1 MLVLLLVLLIA-----VPPRTAEIQ-----LGLREPIGAVSPAFSLTLDAS 44
Qy 87 ITHD-GWLDLSSKRLVTLARLSPAFRFGKRTDFLQFNLRNPARKSRGGPGDYILK 145
Db 45 LARDPRFVALLRHPLKTLASGLSPGLRFGTSTDFLIF----NPNKD-----S 90
Qy 146 NYEDIVRSVDALD-----KQKCKIAQHP-----D 171
Db 91 TWEKVLSEFOAKDVCEAWPSFAVPKLLLTQWPLQELLLAEHSWKHKNTITRSTLD 150
Qy 172 VM-----LELQREKAAQMLHVLKQFSNTYSNLITEPNNYRT 210
Db 151 ILHTPASSGGRFLVGLNALLRRAGLQWDSSNAKOLLYGCAQRSYNISWELGNEPNSFRK 210
Qy 211 MHGRAVNSQLCKDYIQLKSLQIPRIYSRASLYGPNIGRPNKNTIALLDGPMKVGSTV 270
Db 211 KSGICIDGFLGRDPRVHLRQLLSQHPLYRHAELYGLDVGQPKHTQHLRSFMKSGGRAI 270
Qy 271 DAVTHQHCYIDGRVVKVMDFLKRLDLSDOIIRKIQKVNTYTTPGKILWLEGVVTTSAG 330
```

```
Db 271 DSVTHHHYVNGRSATREDFLSPEVLDSFATAIHDVLGIVEATVPGKVKWLGETSAYGG 330
Qy 331 GTNNISDSYAAGFLWNTLGLMANGIDVIRHSFFDHGYNHLVDOONFPLPDYWLSSLY 390
Db 331 GQPLSNTYVAGFMWLDKLGLAARIGDVMVROVSFGAGSYHLVDAGFKPLPDYWLSSLY 390
Qy 391 KRLIGPKVLAVHAGLQRPGRVIRKRLIYAHCTNNHNNYVYRGSITLFIINLHRSR 450
Db 391 KRLVTRVLQASVEQADAR-RP-----RVYLCTNPRHPKYREGDVTLEFALNSNT 441
Qy 451 KKKIAGTLRDLKLVHQLQPYQOGLKSKSVQLNGQPLVMVDDGTLPKLPRLPRAGRT 510
Db 442 QSLQPLKQLWKSVDQYLLPHGKDSILSREVQLNGRLQLQVDDETLPALHEMALPGST 501
Qy 511 LVIPVPTMGFFVVKVNNALAC 531
Db 502 LGLPAFSYGYFVIRNAKIAAC 522

RESULT 9
Q9SDA1 PRELIMINARY; PRT; 521 AA.
AC Q9SDA1
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE HYPOTHETICAL 57.8 KDA PROTEIN.
GN F13G24.30.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Van Der Schueren J., Chuang Y.J., Voet M., Robben J.,
RA Volckaert G., Bancroft I., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133421; CAB62595.1; -
DR InterPro; IPR001254; Trypsin.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 521 AA; 57831 MW; 07D8664A4B305CC2 CRC64;

Query Match 10.8%; Score 302; DB 10; Length 521;
Best Local Similarity 24.8%; Pred. No. 7.7e-16;
Matches 139; Conservative 73; Mismatches 183; Indels 166; Gaps 27;

Qy 75 NENFLSLQLD--PSIIHD-----GWLDFLS----SKRLVTLA-RGLSPAFRFGGR 119
Db 22 DENFVCATLDWNP---HDKCNVDQCWPQSYVINMDLTRPLLTAKAKPLRIRIGSSL 78
Qy 120 TDFLOFQ--NLNRP-----KSRGGCPDYIYLNKYEDDIVRSVDALDKQKCKIAQHPDM 173
Db 79 QDQVYDVGNLKTCPRPFOKMNSG-----LFGFSKGCCKHMKRWD-- 117
Qy 174 LELQREKAAQMLHVL-----LLKEQFSNTYSNLITEPNNYRTMHGRAVN--- 217
Db 118 -EENSEFLTATGAVVIFGLNLRGRHKLKRGKANGGADHINTQDFLNTYSGKVIIDSWEF 176
Qy 218 GSQL-----GKDYIQLKSLQIPRIYSRASLYGPNIGRPNKNTIALDGF-- 262
Db 177 GNELSSGSGVASVAELYGKDLVLKDVIN--KVYKNSWLHPILVAP-----GGFY 226
Qy 263 -----MKVAG-STVDVATVQHCYIDGR-----VVKVMDFLKRLDLSDOIIRKIQ 307
Db 227 EQQWTKLLEISGPSYVDVVT-HHIYNLGSNDPALVKKIMD---PSYLSQVSKTEFDVN 282
```



```
Db 211 -KPLVAPGGFFERQWSELRLSGPGVLDVLTTHIYNLPGNDPKLVNKILD---PNYL 266
QY 297 DTLSDQIRKIOKVNTYTPGKKIWLGVVTTSGAGTNN-----LSDSYAAGFLWLNTLGM 351
Db 267 SGISELFANVNTQIQRCPWAAWV---GEAGGAFNSGGQVSETFINSFWILDQLGI 321
QY 352 LANGIDVVIHRSFFDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHAGLQKRP 411
Db 322 SSKHNTKYICQALVGGFYGLLETFEPNPDYYSALLWHLRMLGKILGVQTTA----- 375
QY 412 PGRVIRKLRVIAHCTNHNHNHNVGSGTFLFIINLHR-----SRKK 452
Db 376 -----SEVLRAVHCXK-----RAGTIIILINLSKHTTFTVAVSNGVKVYLQAESMKR 424
QY 453 IKLAGTLDKLV-----HOYLQYGOEG-LKSKSVOLNGOPLVMVDDGTLP 498
Db 425 KSFLETIKSVKSVGNKASDGYLNREEVHLSF--KQGLRSKIMLLNGKPLVPATGDIP 482
QY 499 ELKPRPLRAGRTLVIPPTMGFFVKKVNALAC 531
Db 483 KLEPVRHGKSPVYINPLSISFVLPTFDAPAC 515
```

## RESULT 12

```
Q9LRC8 PRELIMINARY; PRT; 527 AA.
AC Q9LRC8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BETA-GLUCURONIDASE.
GN SGUS.
OS Scutellaria baicalensis.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Lamiales; Scutellaria.
OX NCBI_TaxID=65409;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20418130; PubMed=10858442;
RA Sasaki K., Taura F., Shoyama Y., Morimoto S.;
RT "Molecular Characterization of a Novel beta-Glucuronidase from
RT Scutellaria baicalensis Georgi.";
RL J. Biol. Chem. 275:27466-27472(2000).
DR EMBL; AB040072; BAA97804.1; -
DR InterPro; IPR001179; FKBP_PPIase.
DR PROSITE; PS00453; FKBP_PPIASE_1; UNKNOWN_1.
SQ SEQUENCE 527 AA; 58772 MW; A5DE7C423F2A1E2B CRC64;
```

```
Query Match 9.2%; Score 256.5; DB 10; Length 527;
Best Local Similarity 24.2%; Pred. No. 3.6e-12;
Matches 129; Conservative 81; Mismatches 207; Indels 115; Gaps 27;
```

```
QY 56 KEKTLILDVSTKNPRTVNENFSLQLD---PSIIHDG---W-----LDFLSSKRLV 102
Db 26 EETTVIKIE---ENPVAQTIDENYCATLDLWPPTKYCNCPWKGSSPLNLD-LNNNIIR 81
QY 103 TLARGLSPAFLRFGGKRTDFLQFNLRNPAKSRGGPGDYVLKKNYEDDIVSDVALDKQK 162
Db 82 NAVKEFAPLKRFGGTQDRLYVQT-----SROEPCDSTFYN-----TNLIDFSH 128
QY 163 GKCTIAQHPDV-----LELQREKAAQMHVLVLLKEQ-----FSN 195
Db 129 ACLSLDRWDEINQFIETGSEAVFGLNALRGKTVKEIGIKDGOYLGTTTAVGEWDYSN 188
QY 196 TYSNLIITEPNNTYTMHGRAVNGSQLG-----KDYI-QLKSLQPIR-IYSRASL 243
Db 189 SKFLIETSLKGYKIRHGWTL-GNELGHTLFIGVSPEDYANDAKKLHELKVEIQDQGT 247
QY 244 YGPNIGRPRKNVIALDGFEMKVAQSTDA-VTWHQHCYIDGRVVKVMDFLKTRLLD-TLSD 301
```

```
Db 248 M-PLIAP--GAIFDLEWYTEFIDRTPELHVATHHMYNLGSGD--DALKOVLLTASFDD 302
QY 302 QIRK-----IOKVNTYTPGKK--IWLEGVVTTSAGGTNNLSDSYAAGFLWLNTLGMLAN 354
Db 303 EATKSMYELQKIVN--RPGTKAVAWIGEAGGAFNSGGQGISNTFINGFWYLNMLGYL 360
QY 355 QGIDVVIHRSFFDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHAGLQKRP 414
Db 361 LDTKFCROTLTGGNYGLLQGTGYIPNPDYYSALLWHLRMLGSKVLKTEIVGPK----- 413
QY 415 VIRDKLRVIAHCTNHNHNHNVGSGTFLFIINLHRSKKIKLA-----GTLRDKLVH 468
Db 414 ----NVYIYAHCAKSN-----GITMLVLN-HDGESSVKISLDESKYGSKRE----EYH 458
QY 469 LQPYQOGLKSKSVOLNGOPLVMVDDGTLPKLPRLRAGRTLVIPPTMGF 520
Db 459 LTPV-NNNLQSELVXLNGELHLHLDPSGVI PALNPVEKDNSKOLEVAPISFMF 509
```

## RESULT 13

```
Q9FZP1 PRELIMINARY; PRT; 536 AA.
AC Q9FZP1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SIMILARITY TO HEPARANASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosomes 5. XI.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB028613; BAB10787.1; -
SQ SEQUENCE 536 AA; 59654 MW; 24202B8E82F3DB0E CRC64;
```

```
Query Match 8.8%; Score 246; DB 10; Length 536;
Best Local Similarity 21.4%; Pred. No. 2.6e-11;
Matches 119; Conservative 86; Mismatches 195; Indels 156; Gaps 25;
```

```
QY 71 VRTVNEFSLQLD---PSIIHDG---WLDFLSSKRL-----VTLARGLSPAFLRFGG 117
Db 41 VGTIDEFICATLDWPPKCDYGCSW-DHASILNLDLTVIVPISKNTAFAPLKRIGG 99
QY 118 KRTDFLQFNLRNPAKSRGGPGDY-----YLKNYEDDIVSDVALDKQCKIAQHPD 171
Db 100 TLQDIIIVET-----PDSKQCLPFTKN-----SSILFGYTGOCPLPMRRD 140
QY 172 VMLEQREKAAQMHVL-----LKE-----QFSNTYSNLIITEPNNY----- 209
Db 141 ELNAPFRKTGTVIFGLNALSGRSIKSNGEATGANNYNAESFIRTAENNYTIDGWELG 200
QY 210 -----TMHGRAVNGSQLGDYIQLKSLQPIRIYGRAS-----LYGPNIGRPRKNV 260
Db 201 NELCGSGVARGVANGQAYADITINLRNIVN--RVYKNVSPMLVIVGP----- 245
QY 261 GPMKY-----AGSTVDVAVTWHQCY-----IDGRVVKVMDFLKTRLLDLSQIRKI 306
Db 246 GFEFVDWTEYLNKAENSLNATT-RHIYDLGPGVDEHLIE--KILNPSYLDQEAQSFSL 302
QY 307 QKVNTYTPGKKIWLGVVTTSG-----GTNNLSDSYAAGFLWLNTLGMLANQGDV 362
Db 303 KNIINKNSTKAWA-----VGESGGAYNSGRNLVSNFVYFWYLDQLGMASLYDFTKYCR 358
QY 363 HSFFDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHAGLQKRPGRVIRDKLRI 422
```

Db	359	QSLIGNYGLLNTNTPNDPYYASALIRQLMGRKALFTTFSGTK	-----KIRS 407
Qy	423	YAHCTHHNNHYVRGSIITLFIINLHRSR---	KKIKIAGTLRDKLVHQVLLQPY-----472
Db	408	YTHCARQSK-----GITVLLNMLDNTTIVAKVELNNSF--	SLRHTKHKMSYKRASSQL 459
Qy	473	--GOEG-----LKSQSVQLNGOPLVWVDGTLPELKPRLRAGRFLVTPP	515
Db	460	FGGPNGVQIREEYHLTAKDNLHLSQTMLENGALQVNSMGDLPPETPIHINSTEPIITAP	519
Qy	516	VTMGFFVVKVNNALAC	531
Db	520	YSIVFVHMNVVVPAC	535
RESULT 14			
Q9VE79			
ID	Q9VE79	PRELIMINARY;	PRT; 935 AA.
CD	Q9VE79		
DT	01-MAY-2000 (TEMBLrel. 13, Created)		
DT	01-MAY-2000 (TEMBLrel. 13, Last sequence update)		
DT	01-MAY-2000 (TEMBLrel. 13, Last annotation update)		
DE	CG14309 PROTEIN.		
GN	CG14309.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_Taxid:7227;		
RP	[1]		
RN	SEQUENCE FROM N.A.		
RC	SPRATN=BERKELEY;		
RC	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Ananastos P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,		
RA	Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,		
RA	Frosier C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,		
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Hartis N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,		
RA	Hoskins D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,		
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,		
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,		
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,		
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,		
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,		
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,		
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,		
RA	Shue B.C., Sidenkianos I., Simpson M., Skupski M.P., Smith T.,		
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,		
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,		
RA	Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,		
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,		
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,		
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,		
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;		
RT	"The genome sequence of Drosophila melanogaster."		
RL	Science 287:2185-2195(2000).		
DR	EMBL; AE003721; AAF55548.1; -		
DR	FLYBase; FBgn0038611; CG14309.		





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:10:18 ; Search time 53.59 Seconds  
(without alignments)  
243.390 Million cell updates/sec

Title: US-09-836-461-2

Perfect score: 2785

Sequence: 1 MRVLCAPFAMPSSNSRPPA.....PVTMGFFVKNVNALACRYR 534

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	936.5	33.6	543	2	US-08-922-170B-10
2	936.5	33.6	543	4	US-09-071-739B-2
3	936.5	33.6	543	4	US-09-260-038B-2
4	932.5	33.5	543	4	US-09-181-336-13
5	889.5	31.9	532	4	US-09-181-336-15
6	805	28.9	380	4	US-09-181-336-19
7	779	28.0	380	4	US-09-181-336-17
8	94.5	3.4	685	2	US-08-878-989-1
9	94.5	3.4	685	3	US-09-136-282-2
10	94.5	3.4	685	4	US-09-272-796-1
11	94.5	3.4	685	4	US-09-505-744-2
12	93.5	3.4	934	1	US-08-215-805A-80
13	93	3.3	471	1	US-08-257-341-9
14	91.5	3.3	915	4	US-09-346-237-2
15	91.5	3.3	928	4	US-09-514-599-4
16	91	3.3	1788	2	US-08-962-284-2
17	89.5	3.2	1118	1	US-07-934-393B-2
18	89.5	3.2	1118	1	US-08-278-089A-2
19	89.5	3.2	1118	2	US-08-838-957A-2
20	89	3.2	632	4	US-09-016-080-1
21	89	3.2	1280	2	US-08-583-276-19
22	89	3.2	1280	6	5206352-4
23	88	3.2	1106	1	US-08-435-675B-5
24	88	3.2	1106	1	US-08-336-257A-8
25	87.5	3.1	752	1	US-08-281-193-2
26	87.5	3.1	752	1	US-08-422-106-2
27	87.5	3.1	752	2	US-08-735-716-2

28	87.5	3.1	752	2	US-08-555-568B-2	Sequence 2, Appli
29	87.5	3.1	752	4	US-09-519-223-2	Sequence 2, Appli
30	87.5	3.1	752	5	PCT-US95-08069-2	Sequence 2, Appli
31	87.5	3.1	834	3	US-08-539-205A-6	Sequence 6, Appli
32	87.5	3.1	1122	1	US-08-278-089A-6	Sequence 6, Appli
33	87.5	3.1	1122	2	US-08-838-957A-6	Sequence 6, Appli
34	87	3.1	1124	1	US-08-323-474-2	Sequence 2, Appli
35	87	3.1	1124	5	PCT-US93-06093-2	Sequence 2, Appli
36	86	3.1	1103	1	US-08-455-543A-53	Sequence 53, Appli
37	86	3.1	1103	2	US-08-223-305C-53	Sequence 53, Appli
38	85	3.1	673	4	US-09-078-347A-2	Sequence 2, Appli
39	85	3.1	1086	6	5386025-8	Patent No. 5386025
40	85	3.1	1280	2	US-08-752-447-2	Sequence 2, Appli
41	84.5	3.0	434	2	US-08-989-925-1	Sequence 1, Appli
42	84.5	3.0	484	3	US-09-080-044-7	Sequence 7, Appli
43	84.5	3.0	533	1	US-08-220-151-15	Sequence 15, Appli
44	84.5	3.0	533	1	US-08-413-118-15	Sequence 15, Appli
45	84.5	3.0	533	3	US-08-473-446-15	Sequence 15, Appli

ALIGNMENTS

RESULT 1  
US-08-922-170B-10  
; Sequence 10, Application US/08922170B  
; Patent No. 5968822  
; GENERAL INFORMATION:  
; APPLICANT: Iris Pecker, Israel Vlodavsky and Elena  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A POLYPEPTIDE  
; TITLE OF INVENTION: HAVING HEPARANASE ACTIVITY AND EXPRESSION OF  
; TITLE OF INVENTION: SAME IN TRANSDUCED CELLS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein  
; STREET: 2940 Birchtree lane  
; CITY: Silver Spring  
; STATE: Maryland  
; COUNTRY: United States of America  
; ZIP: 20906  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk  
; COMPUTER: Twinhead\* Slimnote-890TX  
; OPERATING SYSTEM: MS DOS version 6.2,  
; OPERATING SYSTEM: Windows version 3.11  
; SOFTWARE: Word for Windows version 2.0 converted to  
; SOFTWARE: an ASCII file  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/922.170B  
; FILING DATE: 2 SEP 1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Friedmam, Mark M.  
; REGISTRATION NUMBER: 33,883  
; REFERENCE/DOCKET NUMBER: 910/1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 972-3-5625553  
; TELEFAX: 972-3-5625554  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 543  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-922-170B-10

Query Match 33.6%; Score 936.5; DB 2; Length 543;

Best Local Similarity 37.8%; Pred. No. 3e-91;  
Matches 217; Conservative 78; Mismatches 168; Indels 111; Gaps 12;

Qy 18 PPACIAPGALYALLHLSSLSSQAGDRPLPVDRRAAGLKEKTLILLDLYSTKPNVTVNEN 77  
Db 20 PLGULSPGAL-----PRPA-----QAQDVLDLDFTOEPLHLVSPS 55

Qy 78 FLSLQDPSIHD-GWLDLFLSKRLVTLARGLSPAFLREGGKRTDFLOFQNLNPAKSRG 136  
Db 56 FLSVITANLATDPRFLILLGSPKLTARGLSPAYLREGGKTDFLIF----DPKKEST 111

Qy 137 GPGPDYILKNVEDDVRSDVALDKQKCKIAQ-HPDVMLELQREKAAQHMLVLLKEQF-- 193  
Db 112 PEERSYWSQV 158

Qy 194 ---SNTYS-----NLIL----- 202

Db 159 KFKNTYSRSSVDVLYTFANGSGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGNYI 218

Qy 203 -----TEPNNYRTMHGRAVNSQLGKDYIQLKSLLOPIRIYSRASLYGNIGRPRKNVIA 257  
Db 219 SWELGNEPNSFLKKADIIFINGSQGEDYIQLHKLRLK-STFNKAKLYGPDVQVQVQVQVQVQV 277

Qy 258 LLDGPMKAVGSTVDVATWQHCVIDGRVVKVMDFLKTRLLDLSQDQIRKIQKVNTYTPGK 317  
Db 278 MLKSLFKAGGEVIDSVTHHYYLNGRTATREDFLNPDVLDIFISSVQVQVQVQVQVQVQVQV 337

Qy 318 KIWLEGVVTTSSAGTNNLSDSYAAGFLMNTLGLMLANQIDVVIHRSFDFHGYNHLVDON 377  
Db 338 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMQVFFGAGNYHLVDEN 397

Qy 378 FNPLPDYWLSSLKRLIGPKVLAVHAGLQRPGRVIRDKLRIYAHCTNHHNHYVRG 437  
Db 398 FDPDLYWLSSLKRLIGPKVLAVHAGLQRPGRVIRDKLRIYAHCTNHHNHYVRG 448

Qy 438 SITLFIINLHRSRKKIKLAGTLDKLVHQLQYQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 497  
Db 449 DLTLYAINLHNVTKYLRPLYPFSPNQVDKYLRLPLGPHGLLSKSVQLNGLTLKMYDDQTL 508

Qy 498 PELKPRPLRAGRTLVPVPMGFFVVKVNNALAC 531  
Db 509 PPLMEKPLRPGSSGLPAPFSYFFVIRNAKVAAC 542

RESULT 2  
US-09-071-739B-2  
; Sequence 2, Application US/09071739B  
; Patent No. 6177545  
; GENERAL INFORMATION:  
; APPLICANT: Iris Pecker et al.  
; TITLE OF INVENTION: HEPARANASE SPECIFIC MOLECULAR PROBES  
; TITLE OF INVENTION: AND THEIR USE IN RESEARCH AND MEDICAL  
; TITLE OF INVENTION: APPLICATIONS  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina  
; STREET: 20001 Jefferson Davis Highway, Suite 207  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: United States of America  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk  
; COMPUTER: Twinhead\* Silinote-850TX  
; OPERATING SYSTEM: MS DOS version 6.2  
; OPERATING SYSTEM: Windows version 3.11  
; SOFTWARE: Word for Windows version 2.0 converted to  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071.739B  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/922,180

; FILING DATE: September 2, 1997  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Friedmam, Mark M.  
; REGISTRATION NUMBER: 33,883  
; REFERENCE/DOCKET NUMBER: 910/5  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 972-3-5625553  
; TELEFAX: 972-3-5625554  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 543  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-09-071-739B-2

Query Match 33.6%; Score 936.5; DB 4; Length 543;  
Best Local Similarity 37.8%; Pred. No. 3e-91;  
Matches 217; Conservative 78; Mismatches 168; Indels 111; Gaps 12;

Qy 18 PPACIAPGALYALLHLSSLSSQAGDRPLPVDRRAAGLKEKTLILLDLYSTKPNVTVNEN 77  
Db 20 PLGULSPGAL-----PRPA-----QAQDVLDLDFTOEPLHLVSPS 55

Qy 78 FLSLQDPSIHD-GWLDLFLSKRLVTLARGLSPAFLREGGKRTDFLOFQNLNPAKSRG 136  
Db 56 FLSVITANLATDPRFLILLGSPKLTARGLSPAYLREGGKTDFLIF----DPKKEST 111

Qy 137 GPGPDYILKNVEDDVRSDVALDKQKCKIAQ-HPDVMLELQREKAAQHMLVLLKEQF-- 193  
Db 112 PEERSYWSQV 158

Qy 194 ---SNTYS-----NLIL----- 202

Db 159 KFKNTYSRSSVDVLYTFANGSGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGNYI 218

Qy 203 -----TEPNNYRTMHGRAVNSQLGKDYIQLKSLLOPIRIYSRASLYGNIGRPRKNVIA 257  
Db 219 SWELGNEPNSFLKKADIIFINGSQGEDYIQLHKLRLK-STFNKAKLYGPDVQVQVQVQVQVQV 277

Qy 258 LLDGPMKAVGSTVDVATWQHCVIDGRVVKVMDFLKTRLLDLSQDQIRKIQKVNTYTPGK 317  
Db 278 MLKSLFKAGGEVIDSVTHHYYLNGRTATREDFLNPDVLDIFISSVQVQVQVQVQVQVQVQV 337

Qy 318 KIWLEGVVTTSSAGTNNLSDSYAAGFLMNTLGLMLANQIDVVIHRSFDFHGYNHLVDON 377  
Db 338 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMQVFFGAGNYHLVDEN 397

Qy 378 FNPLPDYWLSSLKRLIGPKVLAVHAGLQRPGRVIRDKLRIYAHCTNHHNHYVRG 437  
Db 398 FDPDLYWLSSLKRLIGPKVLAVHAGLQRPGRVIRDKLRIYAHCTNHHNHYVRG 448

Qy 438 SITLFIINLHRSRKKIKLAGTLDKLVHQLQYQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 497  
Db 449 DLTLYAINLHNVTKYLRPLYPFSPNQVDKYLRLPLGPHGLLSKSVQLNGLTLKMYDDQTL 508

Qy 498 PELKPRPLRAGRTLVPVPMGFFVVKVNNALAC 531  
Db 509 PPLMEKPLRPGSSGLPAPFSYFFVIRNAKVAAC 542

RESULT 3  
US-09-260-038B-2  
; Sequence 2, Application US/09260038B  
; Patent No. 6348344  
; GENERAL INFORMATION:  
; APPLICANT: Maty Ayal-Hershkovitz et al.  
; TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR  
; EXPRESSING RECOMBINANT HEPARANASE



Db 338 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 397  
QY 378 FNPLPDYWLSLLYKRLIGPKVLAVHVGLOKRPGRVIROKLRIYAHCTNHHNNHYVRG 437  
Db 398 FDPPLDYWLSLLFKLVGTKVLMAVSGVSKRR-----KLRVYLHCTNTDNPRIYKEG 448  
QY 438 STTFTIINLHRSRKKIKLAGTLRDLKHQYLLQPYGOGLSKSVQNLGQPLVWVDDGTL 497  
Db 449 DLTVALNHNVTYKLRPLPYPSNKQVDKYLRLPLGPHGLLSKSVQNLGLTLKMWDDQTL 508  
QY 498 PELKPRPLRAGRTL 511  
Db 509 PPLMEKPLRPGSSL 522

RESULT 5  
US-09-181-336-15  
; Sequence 15, Application US/09181336A  
; Patent No. 6242238  
; GENERAL INFORMATION:  
; APPLICANT: FREEMAN, Craig Geoffrey  
; APPLICANT: PARISH, Richard  
; APPLICANT: HAMDORF, Brenton James  
; APPLICANT: HULETT, Mark Darren  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN  
; FILE REFERENCE: 032505-004  
; CURRENT APPLICATION NUMBER: US/09/181,336A  
; CURRENT FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: AU PP0062  
; EARLIER FILING DATE: 1997-10-28  
; EARLIER APPLICATION NUMBER: AU PP0812  
; EARLIER FILING DATE: 1997-12-09  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 15  
; LENGTH: 532  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-181-336-15

Query Match 31.9%; Score 889.5; DB 4; Length 532;  
Best Local Similarity 37.7%; Pred. No. 3.1e-86;  
Matches 209; Conservative 75; Mismatches 159; Indels 111; Gaps 12;

QY 18 PPACIAPGALYLALLHLSSLGQADRRPLPVDRAAGLKEKTLILLDVSTKPNVTVNEN 77  
Db 20 PLGPLSPGAL-----PRPA-----QAQDVVDLDFFTQBPGLHVSPTS 55  
QY 78 FLSLQDPSI IHD-GWLDLFLSKRLVTLARGLSAPFLREGGKRTDFLOFONLRNPAKSRG 136  
Db 56 FLUSVTDANLADPREFLIILGSPKLTARGLSAPFLREGGKTDFLI-----DPKKEST 111  
QY 137 GPGPDYLLKNYEDDIRSDVALDKOKGCKIAQ-HPDVMLELQREAAQMHVLLKKEQF-- 193  
Db 112 FEERSYQSOVNQDI-----CRYGSIPDPVEEKLRLWPYQEQQL-LLREHYQK 158  
QY 194 ---SNYS-----NLIL----- 202  
Db 159 KFNKSTYSRSSVDLYTFANCGLDLIFGLNALLRTADLQWNSSNAQLLDLYCSSKGYNI 218  
QY 203 -----TEPNYRTMHGRAVNSQLGKDYIQLKSLQPIRIYSRASLYGNIGRPRKNVIA 257  
Db 219 SWELGNEPNSFLKADIFINGOLGEDFIQLHKLRR-STFNKAKLYGPDVGQPRKKTAK 277  
QY 258 LLDGFMKAVAGSTVDATWQHCHYIDGRVVKVMDFLKTRLDLTSDQIRKIQKVNTYTPGK 317  
Db 278 MLKSFLLKAGGEVIDSVTWHYLYNGRATREDFLNPDVLDIFSSQKVFQVVESTPCPK 337  
QY 318 KIWLEGVVTTSAGTNNLSDSYAAGFLMNTLGMLANOGIDVVIHRSFFDHGNYHLVDQ 377

Db 338 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 397  
QY 378 FNPLPDYWLSLLYKRLIGPKVLAVHVGLOKRPGRVIROKLRIYAHCTNHHNNHYVRG 437  
Db 398 FDPPLDYWLSLLFKLVGTKVLMAVSGVSKRR-----KLRVYLHCTNTDNPRIYKEG 448  
QY 438 STTFTIINLHRSRKKIKLAGTLRDLKHQYLLQPYGOGLSKSVQNLGQPLVWVDDGTL 497  
Db 449 DLTVALNHNVTYKLRPLPYPSNKQVDKYLRLPLGPHGLLSKSVQNLGLTLKMWDDQTL 508  
QY 498 PELKPRPLRAGRTL 511  
Db 509 PPLMEKPLRPGSSL 522

RESULT 6  
US-09-181-336-19  
; Sequence 19, Application US/09181336A  
; Patent No. 6242238  
; GENERAL INFORMATION:  
; APPLICANT: FREEMAN, Craig Geoffrey  
; APPLICANT: PARISH, Richard  
; APPLICANT: HAMDORF, Brenton James  
; APPLICANT: HULETT, Mark Darren  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN  
; FILE REFERENCE: 032505-004  
; CURRENT APPLICATION NUMBER: US/09/181,336A  
; CURRENT FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: AU PP0062  
; EARLIER FILING DATE: 1997-10-28  
; EARLIER APPLICATION NUMBER: AU PP0812  
; EARLIER FILING DATE: 1997-12-09  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Rattus sp.  
US-09-181-336-19

Query Match 28.9%; Score 805; DB 4; Length 380;  
Best Local Similarity 47.3%; Pred. No. 1.9e-77;  
Matches 155; Conservative 57; Mismatches 106; Indels 10; Gaps 2;

QY 204 EPNNYRTMHGRAVNSQLGKDYIQLKSLQPIRIYSRASLYGNIGRPRKNVIALLDGPM 263  
Db 62 EPNSFWKKAHISIDGLQGEFVELHKLQK-SAFQNAKLYGPDIGQPGKGVKLLRSEFL 120  
QY 264 KVAGSTVDATWQHCHYIDGRVVKVMDFLKTRLDLTSDQIRKIQKVNTYTPGKKIWLGE 323  
Db 121 KAGGEVIDSLTWHYLYNGRVATKEDFLSSDVLDTFILSVQILKVTKEMTGKKVWLGE 180  
QY 324 VYTTGAGTNNLSDSYAAGFLMNTLGMLANOGIDVVIHRSFFDHGNYHLVDQNFNPLPD 383  
Db 181 TSSAYGGAPLLSDTFAAGFMWLDKLGSAQLGIEVVMRQVFFGAGNYHLVDENFEPLPD 240  
QY 384 YWLSLLYKRLIGPKVLAVHVGLOKRPGRVIROKLRIYAHCTNHHNNHYVRGSTITFLI 443  
Db 241 YWLSLLFKLVGPKVLMRSVKGPD-----RKLRLYLHCTNHYHPRYREGDUTLV 291  
QY 444 INLHRSRKKIKLAGTLRDLKHQYLLQPYGOGLSKSVQNLGQPLVWVDDGTLPELKPR 503  
Db 292 LNLHNVTYKLRPLPMPFSPVDKYLKLPFGSDGLLSKSVQNLGLTLKMWDEQTLPALTEK 351  
QY 504 PLRAGRTIVIPVTWGTGFFVVKVNNALAC 531  
Db 352 PLPAGSSLSVPAFSYGFFVIRNAKTAAC 379

RESULT 7  
US-09-181-336-17

; Sequence 17, Application US/09181336A  
; Patent No. 6242238  
; GENERAL INFORMATION:  
; APPLICANT: FREEMAN, Craig Geoffrey  
; APPLICANT: PARISH, Richard  
; APPLICANT: HAMDORE, Brenton James  
; APPLICANT: HULETT, Mark Darren  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN  
; FILE OF INVENTION: ENDOLUCURONIDASE AND USES THEREFOR  
; FILE REFERENCE: 032505-004  
; CURRENT APPLICATION NUMBER: US/09/181,336A  
; CURRENT FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: AU PP0062  
; EARLIER FILING DATE: 1997-10-28  
; EARLIER APPLICATION NUMBER: AU PP0812  
; EARLIER FILING DATE: 1997-12-09  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; US-09-181-336-17

Query Match 28.0%; Score 779; DB 4; Length 380;  
Best Local Similarity 46.3%; Pred. No. 1.2e-74;  
Matches 152; Conservative 56; Mismatches 110; Indels 10; Gaps 2;

QY 204 EPNVYTMHGRAVNSOLGCKDYIOLKSLLOPIRIYSRSLYGNIGRPRKNVIALLDGPM 263  
DB 62 EPNFVKKAHIIIDQLGDEFEVHKLQQR -SAFONAKLYGPDIGOPRGKTVKLLRSL 120  
QY 264 KVAGSTVDVAVTWHQYIDGRVYKVMDFLTKRLDITLSDQIRKIQKVVNTYTPGKKIWLGE 323  
DB 121 KAGGEVIDSLTWHYLYNGRIATKEDFLSSDVLDTILSVOKILKYTKETPGKKVWLGE 180  
QY 324 VVTSAGGNNLSDSVAGFLMLNTGLMLANOGIDVIRHSFFDHGYNHLVDQNFPLPD 383  
DB 181 TSSAYGGGAPLNSNTPAAGFMFLDKLGLSAQMGIEVVRQVFFGAGNYHLVDNFEPFLPD 240  
QY 384 YWLSLLYKRLIGPKVLAVHAGLORPRGVRIRDKRLRIYAHCTNNHHNHYVYRGSTLFI 443  
DB 241 YWLSLLFKLVGPRVLLSRVKGPD-----RSKRLYLHCTNYHYRPOEGDITLYV 291  
QY 444 INLHRSRKKIKLAGTLRDLKLVHQLQPYQGEGLKSKSVQLNGOPLVMVDDGTLPELKPR 503  
DB 292 LNLHNVTKHLKVPPLFRKPPVDFYLLKPSGPDGLLSKSVQLNGOILKMWDEQTLPALTEK 351  
QY 504 PLRAGRTIVPPVTMGFFVVKVNNALAC 531  
DB 352 PLPAGSALSPLAFSYGFFVIRDAKTAAC 379

RESULT 8  
US-08-878-989-1  
; Sequence 1, Application US/08878989  
; Patent No. 5885803  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Guegler, Karl G.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN  
; NUMBER OF INVENTION: KINASES  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto

; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/878,989  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0321 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 685 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: HUVENOB01  
; CLONE: 39043  
; US-08-878-989-1

Query Match 3.4%; Score 94.5; DB 2; Length 685;  
Best Local Similarity 23.1%; Pred. No. 0.56;  
Matches 84; Conservative 46; Mismatches 144; Indels 89; Gaps 18;

QY 129 RNPASRGSGPGPDYLYLKNYEDDIVRSDVLDKQCKIAQHDPVMLELOREKAAQ----- 183  
DB 84 RGVLGKGG-----FAKCYE-----MTDLTNKVAAKIIPHSRVAKPHQREKIDKEIELH 134  
QY 184 -----MHLVLLKQFNTYSNLITPNNVYTMHGRAVNSOLGKDYIOLKSLLOP-IRI 237  
DB 135 RILHHKVVQFYHYFEDKENIYILLEVCSRRSM-----AHILK---ARKVLTEPEVRY 184  
QY 238 YSRASLYGNIGRPRKNVIALLDGPMKVGASTVDVAVTWHQYIDGRVYKVMDFLTKRLTD 297  
DB 185 YLRQIVSGLKYLHEQE-----ILHRDLKLGNNFFIN-----EAMELKVGDFGLAARLE 231  
QY 298 TLDQIRKIQKVVNTYTP-----GKKIWLGEVW---TTSAG-----GTNNLSDSY-- 339  
DB 232 PLEHRRRTICGTNYLSPEVLNKGHCESDIWALGCVMTMLLGRPPFTTNLKETYRC 291  
QY 340 --AAGFLWLNTL-----GMLANQ-----GIDVYIRHSFFDHGY--NHLVDQNFNPL 381  
DB 292 IREARYTMPSSLAPAKHLIASMLSKNPEDRPSLDIIRHDFLQGFPTDRLSSSCCHTV 351  
QY 382 PDYWLSLLYKRLIGPKVLAVHAGLORKPR-----PGRVIRDKRLRIYAHCTNNHHNHYVRC 437  
DB 352 PDFHLSPAKNFF-KKAAALFGGKKDKARYIDTHNRVSKEDDIY-----KLRLHDLKKT 405  
QY 438 SIT 440  
DB 406 SIT 408

RESULT 9  
US-09-136-2-2  
; Sequence 2, Application US/09136282  
; Patent No. 6063609  
; GENERAL INFORMATION:

APPLICANT: ANDERSON, KAREN  
APPLICANT: JACKSON, JEFFREY  
APPLICANT: HANSBURY, MICHAEL  
APPLICANT: NERURKAR, SANDHYA  
APPLICANT: ROZHAK, AMY  
APPLICANT: BOUZYK, MARK  
TITLE OF INVENTION: HUMAN SERUM INDUCIBLE KINASE (SNK)  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ratner & Prestia  
STREET: P.O. Box 980  
CITY: Valley Forge  
STATE: PA  
COUNTRY: USA  
ZIP: 19482  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/136,282  
FILING DATE: 20-AUG-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/056,112  
FILING DATE: 20-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Prestia, Paul F  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: GH-70231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700  
TELEFAX: 610-407-0700  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 685 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-136-282-2

Query Match 3.4%; Score 94.5; DB 3; Length 685;  
Best Local Similarity 23.1%; Pred. No. 0.56;  
Matches 84; Conservative 46; Mismatches 144; Indels 89; Gaps 18;  
QY 129 RNPASRGPGPDYILKNYEDDIVRSVDALDKQCKIAQHDPVDMLELQREKAAQ-----183  
DB 84 RGVVLKGG-----FACYE---MTDLTNKVYAAKIIPHSRVAKPHQREKIDKTELH 134  
QY 184 -----MHLVLLKQFSWTYNSNLTPEPNYRTVHGRAVNSQLGKDYIQLKSLQP-IRI 237  
DB 135 RILHKKHVQVQHYFYFEDKENIYILLEYCSRRSM-----AHILK---ARKVLTEPEVRY 184  
QY 238 YSRASLYGPNIGRPRKNVIALDGFMKVAGSTVDVAVTQHCYIDGRVVKYMDFLKTRLLD 297  
DB 185 YLRQIVSGLKYLHEQE-----ILHRDLKLGNNFFN-----EAMELKVDGFLGAARLE 231  
QY 298 TLDQIRKIQKVVNTYTP-----GKKIWLEGVY--TTSAG-----GTNNLSDSY-- 339  
DB 232 PLEHRRRTICGTPNYSPEVLNKGHCESDIWALGCVMTMLLGRPPFTTNLKETIRC 291  
QY 340 --AAGFLWLNTL-----GMLAQ-----GIDVVIRISFFDHGY--NHLVDQNFNPL 381  
DB 292 IREARYTMPSSLAPAKHLIASMLSKNPEDRPSLDDIIRHDFLQGTDPRLSSCOCHT 351  
QY 382 PDVWLSLLYKRLIGPKVLAVHAGLQKRP-----PGRVIRDKLRIYAHCTNHHNNHYVRG 437  
DB 352 PDFHLSPPAKNFF-KKAAALFGGKKDKARYIDTHNVSKEDEDIY-----KLHRDLKKT 405

QY 438 SIT 440  
DB 406 SIT 408  
RESULT 10  
US-09-272-796-1  
; Sequence 1, Application US/09272796  
; Patent No. 6207148  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Guegler, Karl G.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Goli, Suriya K.  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN  
; TITLE OF INVENTION: KINASES  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/272,796  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/878,989  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0321 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 685 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: HUVENOB01  
; CLONE: 39043  
; US-09-272-796-1

Query Match 3.4%; Score 94.5; DB 4; Length 685;  
Best Local Similarity 23.1%; Pred. No. 0.56;  
Matches 84; Conservative 46; Mismatches 144; Indels 89; Gaps 18;  
QY 129 RNPASRGPGPDYILKNYEDDIVRSVDALDKQCKIAQHDPVDMLELQREKAAQ-----183  
DB 84 RGVVLKGG-----FACYE---MTDLTNKVYAAKIIPHSRVAKPHQREKIDKTELH 134  
QY 184 -----MHLVLLKQFSWTYNSNLTPEPNYRTVHGRAVNSQLGKDYIQLKSLQP-IRI 237  
DB 135 RILHKKHVQVQHYFYFEDKENIYILLEYCSRRSM-----AHILK---ARKVLTEPEVRY 184  
QY 238 YSRASLYGPNIGRPRKNVIALDGFMKVAGSTVDVAVTQHCYIDGRVVKYMDFLKTRLLD 297





Db 190 DTSEIQISQGLTKLVNQVGLGTGLGDKLNFSQFSKAGLGLVVISGLLSGATAALVLADKN 249

QY 340 A-----AGFLWNLNTGLMANQ 355

Db 250 ASTDRKVGAGF-----ELANQ 265

RESULT 13

US-08-257-341-9

; Sequence 9, Application US/08257341

; Patent No. 5525491

; GENERAL INFORMATION:

; APPLICANT: HUSTON, JAMES S

; APPLICANT: OPPERMANN, HERMANN

; APPLICANT: TIMASHEFF, SERGE N

; TITLE OF INVENTION: SERINE RICH PEPTIDE LINKER

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CREATIVE BIOMOLECULES, INC./PATENT DEPT.

; STREET: 35 SOUTH STREET

; CITY: HOPKINTON

; STATE: MA

; COUNTRY: USA

; ZIP: 01748

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/257,341

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION NUMBER: US/07/842,149

; APPLICATION DATA:

; FILING DATE:

; APPLICATION NUMBER: US 07/662,226

; FILING DATE: 27-FEB-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: CAMPBELL ESQ, PAULA A

; REGISTRATION NUMBER: 32,503

; REFERENCE/DOCKET NUMBER: CRP-064CP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/248-7000 (ATTY)

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 471 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-257-341-9

Query Match 3.3%; Score 93; DB 1; Length 471;

Best Local Similarity 19.2%; Pred. No. 0.43;

Matches 90; Conservative 71; Mismatches 152; Indels 156; Gaps 21;

QY 55 LKSKTLLDVSTKNPVRTVNEFLSLQDPSIIHDGWLDFLSSKRLVTLARGLS----- 109

Db 68 LKDRINIVLSRELKEPR--GAHFLAKSLDDLRLLEQPELASKVDMVWVGSSVYQEA 125

QY 110 ---PAFLRFGGKRTDFLQFONLRNPAKSRGGPGDYLYLKNYEDIVRSVDALDKRGCKI 166

Db 126 MNQPGHLRLVTR-----IMQEFESDTFFPEIDLKGYK--LL 160

QY 167 AQHPDVMLELQREKAAQMHVLVLLKEQFSNTYSNLIILTEPNNYTMHGRAVNSQLGKDYI 226

Db 161 PEYPGVLSEVOEEKGIKYKEFVEYKDKAS-----VTVSSSGSSSSSSSSSSSMIEOD-- 212

QY 227 QLSKLLQPIRIYSRASLYGPNIGRPNKRNIALL--DGFMKVGASTVDVATWQHCYIDGRV 284

Db 213 -----GLHAGSPAAWVERLFGYDWAQQTICGS--DAAVFR--LSAGGRP 252

QY 285 VKVMDFLKTRLLDTSQIRKIQKVVNTYTPGKKIWLEGVVTTSGGTTNNLSDSYAAGFL 344

Db 253 VL---FVKTDL-----SGALNELQDE--AARLS 275

QY 345 WLNTGLMANOGIDVIRHSPFDHGYNHLVDQNFNPLPDYWLSLLYKRLGPKVLAVHVA 404

Db 276 WLATTGVPCAAVLDDVT-----EAGRD-----W--LLLGEVPGQDLSLHLA 315

QY 405 GLQRKPRGRVIRDKLRIY-----AHCTNHHNHNVVRGSIITLFIINLHRSRKKIKLAGTL 459

Db 316 PAEKY-----SIMADAMRLHLTDPATCFDHOAKH-----RIERATRWEAGLYD 361

QY 460 RDKLVHVOYLLQPYGOEGL-----KSKSVQLNGQPLVMV--DDGTLPEL 500

Db 362 QDDLDEH-----OGLAPAEFLARLKARMPDGEDLVVTHGDACLPI 403

RESULT 14

US-09-346-237-2

; Sequence 2, Application US/09346237A

; Patent No. 6265197

; GENERAL INFORMATION:

; APPLICANT: Bisgard-Frantzen, Henrik

; APPLICANT: Svendsen, Allan

; TITLE OF INVENTION: Starch Debranching Enzymes

; FILE REFERENCE: 5629.200-US

; CURRENT APPLICATION NUMBER: US/09/346,237A

; CURRENT FILING DATE: 1999-07-01

; EARLIER APPLICATION NUMBER: PA 1998 00868

; EARLIER FILING DATE: 1998-07-02

; EARLIER APPLICATION NUMBER: 60/094,353

; EARLIER FILING DATE: 1998-07-28

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 915

; TYPE: PNT

; ORGANISM: Bacillus deramificans

; FEATURE:

; NAME/KEY: PEPTIDE

; LOCATION: (1)-(915)

; OTHER INFORMATION: Pullulanase

US-09-346-237-2

Query Match 3.3%; Score 91.5; DB 4; Length 915;

Best Local Similarity 23.0%; Pred. No. 1.9;

Matches 79; Conservative 38; Mismatches 127; Indels 99; Gaps 19;

QY 43 DRRLPVDRAAGLKE--KTLILLDVSTKNP--VRTVNEFLSLQDPSIIHDGWLDFLSS 98

Db 424 DVRDFSIDPNSGMKNKGYLALTEKGTGPDNVKT-----GID--SL 463

QY 99 KRLVTLARGLSPAFLRFGGKRTDFLQFONLRNPAKSRGGPGDYLYLKNYEDDIVRSVAL 158

Db 464 KOLGTHVOLMVFASNSVDETFD-----PTDNMGYDP-----RNY--DVPEGQYAT 508

QY 159 DRQKCKIAQHPDVMLELQREKAAQMHVLVLLKEQFSNTYSNLIILTEPN--NYRTHM--GRAV 216

Db 509 NANGNARIKEFEKEMVLSLHREHIGVMVYVNHTEFATQISDFDKIVPEYYRTDDAGNYT 568

QY 217 NGSQGLKDYIQLSKLLQPIRIYSRASLYGPNIGRPNKRNIALLDGFMKVGASTVDVAT--W 275

Db 569 NGSGTGNETAERPMVQKF-----IIDSLEYW 595

QY 276 QHCY-IDGRVVKVMDPLKTRLLDTSQIRKIQKVVNTYTPGKKIWLEGVVTTSGGTTNN 334

Db 596 VNEYHIDGRFDMALLGK---DTMS-----KAASELHAINPG--ITALYGEPT--GGTSA 644

QY 335 LSDSYAAGFLWNLNTGLMANQGDIVVIRHSPFDHGYNHLVDQN 377

Db 645 LPDDQ-----LLTKG--AOKGMGAV-----FNDNLRLNLDGN 675

Search completed: July 30, 2002, 08:15:01  
Job time: 283 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:08:23 ; Search time 141.47 Seconds  
(without alignments)  
419.265 Million cell updates/sec

Title: US-09-836-461-2  
Perfect score: 2785  
Sequence: 1 MRVLCAPPEAMPSSNSRPPA.....PVTMGFFVKNVNALACRYR 534

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2785	100.0	534	22	AAB85216
2	2746	98.6	592	22	AAU07424
3	2746	98.6	592	22	AAB85215
4	2746	98.6	592	22	AAU07424
5	2736	98.2	592	22	AAU07424
6	2564	92.1	492	22	AAB81062
7	2485	89.2	538	22	AAU07418
8	2478	89.0	480	22	AAU07418
9	2478	89.0	480	22	AAB85217
10	2478	89.0	480	22	AAU07423
11	2270	81.5	439	22	AAU07423

12	1314	47.2	262	22	AAM24147	Human EST encoded
13	1112	39.9	214	22	AAM99905	Human excretory re
14	1112	39.9	214	22	AAM43704	Human bladder anti
15	944.5	33.9	535	21	AAB08851	A murine heparanas
16	936.5	33.6	543	20	AAU02345	A human heparanas
17	936.5	33.6	543	21	AAB08849	Amino acid sequenc
18	936.5	33.6	543	21	AAU57590	Human heparanas
19	936.5	33.6	543	21	AAU52990	Human heparanas
20	936.5	33.6	543	22	AAU97635	Human heparanas
21	936.5	33.6	592	20	AAU02346	A human heparanas
22	936.5	33.6	592	21	AAB08850	Amino acid sequenc
23	933.5	33.5	543	22	AAB88361	Human membrane or
24	932.5	33.5	543	20	AAU17082	Human heparanas
25	932.5	33.5	543	22	AAB86206	Human heparanas
26	932.5	33.5	588	20	AAU30124	A human protein wi
27	929.5	33.4	530	20	AAU34173	Human pre-prohepar
28	889.5	31.9	532	20	AAU17083	Seq ID No: 15 Of W
29	805	28.9	380	20	AAU17085	Rat heparanas enz
30	779	28.0	380	20	AAU17084	Mouse heparanas
31	778	27.9	156	22	AAG65963	Human heparanas-1
32	468.5	16.8	488	22	AAB31469	Amino acid sequenc
33	440.5	15.8	488	22	AAB31472	Amino acid sequenc
34	439.5	15.8	488	22	AAB31470	Amino acid sequenc
35	422.5	15.2	488	22	AAB31471	Amino acid sequenc
36	278	10.0	137	22	AAG65961	Human heparanas-1
37	278	10.0	159	22	AAG65964	Human heparanas-1
38	270	9.7	112	22	AAU07425	Human heparanas-1
39	247	8.9	256	21	AAG13479	Arabidopsis thalia
40	203	7.3	118	22	AAG65962	Human heparanas-1
41	171	6.1	935	22	ABU69219	Drosophila melanog
42	162	5.8	115	22	AAB85218	Mouse heparanas-1
43	144	5.2	38	20	AAU34186	Human pre-prohepar
44	127	4.6	24	22	AAB85220	Heparanas-like pr
45	127	4.6	25	22	AAB85221	Heparanas-like pr

ALIGNMENTS

RESULT 1  
AAB85216  
ID AAB85216 standard; Protein; 534 AA.

XX	AC	AAB85216;
XX	AC	AAB85216;
XX	DT	07-SEP-2001 (first entry)
XX	DE	Heparanas-like protein Hpa2 splice variant #2.
XX	DE	Heparanas-like protein Hpa2 splice variant #2.
XX	KW	Heparanas; splice variant; homologue; heparanas-like protein; Hpa2;
XX	KW	cytostatic; neuroprotective; cerebroprotective; immunosuppressive;
XX	KW	antipsoriatic; nontropic; antiinflammatory; antiarthritic; antiasthmatic;
XX	KW	antidiabetic; antiarteriosclerotic; vulnerary.
OS	OS	Homo sapiens.
XX	XX	
XX	PN	WO200146392-A2.
XX	PD	28-JUN-2001.
XX	XX	
XX	PF	21-DEC-2000; 2000WO-GB04963.
XX	PR	22-DEC-1999; 99GB-0030392.
XX	PR	07-APR-2000; 2000GB-0008713.
XX	PA	(OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX	PI	McKenzie EA, Stamps AC, Terrett JA, Tyson KL;
XX	DR	WPI: 2001-418056/44.
XX	DR	N-PSDB; AAB22672.
XX	PT	Novel homologs of heparanas, present in three splice variants, useful

→ post dated

for identifying agents that modulate heparanase, useful in the treatment and/or prophylaxis of abnormal levels of heparanase

treatment and/or prophylaxis of abnormal levels of heparanase

Claim 1: Fig 2: 97pp; English.

The invention provides a homologue to heparanase which is present in three splice variants. The heparanase homologue polypeptide is useful in the treatment of a human or non-human animal or for use in diagnosis. Vectors comprising the heparanase homologue polynucleotides are useful in the transformation or transfection of a prokaryotic or eukaryotic host. The modulators of the polypeptide are useful in the manufacture of a medicament for the treatment and/or prophylaxis of a condition/disease associated with abnormal levels of the heparanase homologue, including cancer, central nervous system (CNS) and neurodegenerative diseases, cardiovascular diseases such as stenosis following angioplasty and atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus, allografts, inflammatory diseases, arthritis, vascular stenosis, tumour growth and progression, asthma, Alzheimer's disease, diabetic retinopathy, wound healing and inflammation. The polypeptide is also useful in diagnosis and research. The present sequence represents the amino acid sequence of the mid-sized splice variant of the heparanase-like protein Hpa2 of the invention.

AA	Sequence	534 AA;
SO		

```
Query Match      100.0%; Score 2785; DB 22; Length 534;
Best Local Similarity 100.0%; Pred. No. 8.3e-274;
Matches 534; Conservative 0; Mismatches 0; Indels 0;
```

Qy	1	MRVLCAPPEAMPSNSRPPACIAPGALYALALLHLSLSSOAGDRRPLPVDRAAGLKEKTL	60
Db	1	mrvlcapseamspnsrppacIapgalYalallhlslssagagorrlpdydraagIkektl	60
Qy	61	ILLDVSTKNPVRVNEFNLSLOLDPSIIHDGWLDFLSSKRLVTLARGLSPAFLRFGGKRT	120
Db	61	illdvstknprvtnefnlsldpsIihdgwldflsskrlvtlarglsPafIrfggkrt	120
Qy	121	DFLOFQMLRNPAKSRGGPGDPDYLYKNVEDDIVRSDVALDKQCKCKTAQHPDVMLELOREK	180
Db	121	dflqfqlnrnpaksr9gpgpdylyknyeddivrsdvaldkqgckIaqdpvmleIgrek	180
Qy	181	AAQMHVLVLLKEQFSNTYSNLITEPNNYRPMHGRAVNGSOLGXDYTLQLKSLLQPIRIYSR	240
Db	181	aaqmhlvlllkeqfsntySnlITepnnYrPMhgravngsQlgxdYtlqLksllqPiriYsr	240
Qy	241	ASLYGPNIGRPRKNVITALLDGFEMKVAGSTVDATWQHICYIDGRVVKVMDFLTKRLTDLTIS	300
Db	241	aslygpnigrprknvIaLLdGfemkvagstvdAtwqhicyIdgrvVkvMdfLtkrltDltIs	300
Qy	301	DQIRKIQKVNTYTPGKKIWLEGVWVITSAGTNNLSDSYAAGFLWTLTGLMLANOGIDVV	360
Db	301	dqIrKIdkvntytPgkKIwlegvvtTsaggtNNlSDsYaaGflwTlTglmlanogIdvv	360
Qy	361	IRHSFFDHGYNHLVDQNFNLPDYLWLSLLYKRLIGPKVLAVHVAGLQRPKRPGRVIRDKL	420
Db	361	IrhsffdhgynhlvDqnfNlpdYlWlsllYkrlgPkvlavhVagLqrKpRpgrvIrDkl	420
Qy	421	RIYAHCTNHHNNHVVRGSITLFIINLHRSRKKIKLAGTLRDKLVHQYLLQPYQGEGLKSK	480
Db	421	riyahctnhnnhmyvrGsItlfiInlHrsrKkIagtlrDkLvHqYllqPyqgeglKsk	480
Qy	481	SVQLNGOPLVMVWDGTLPELKPRLPRAGRTLIVIPPTMGFFVVKVNNVALACRYR	534
Db	481	svqlngoplvmvwdGtlPeLkpRlPrLagrtLIVipPtMGffvVkvNnValacryr	534

RESULT 2  
AAU07424  
ID: AAU07424 standard: Protein: 592 AA.

XX  
AC AAU07424:

XX

18-DEC-2001	(first entry)
Human	heparanase-like protein splice variant #1.
Human;	Immunosuppressive; antiarthritic; antirheumatic; cytostatic; antiproliferative; cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide; fungicide; ophthalmological; extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis; hyperproliferative disorder; neoplasm; cardiovascular disorder; cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection; nervous system disorder; Alzheimer's disease; ocular disorder; sunburn; wound healing; food additive; heparanase.
HOmo sapiens.	
WO200179253-A1.	
25-OCT-2001.	
11-APR-2001;	2001WO-US11643.
18-APR-2000;	2000US-198123P.
(HUMA-)	HUMAN GENOME SCI INC.
Fiscella M, Shi Y, Ebner R, Ruben SM;	
WPI; 2001-611720/70.	
N-PSDB; AAS13848.	
New nucleic acids encoding extracellular matrix polypeptides, for diagnosing, treating, preventing or ameliorating human disorders and disease, such as, autoimmune, hyperproliferative or cardiovascular disorders -	
Disclosure; Page 14; 308pp; English.	
The invention relates to novel isolated polynucleotides (I) encoding extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by (I) are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. The antibodies to the polypeptides can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis.	
hyperproliferative disorders e.g. neoplasms of the breast or liver. cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. The present sequence represents the amino acid sequence of human heparanase-like protein, splice variant #1.	
Sequence	592 AA;
Query Match	98.6%; Score 2746; DB 22; Length 592;
Best Local Similarity	90.2%; Pred. No. 9.1e-270;
Matches 534.	Conservative 0; Mismatches 0; Indels 58; Gaps

RESULT 2

AAU07424

—ET—

XX

AC

Db 61 illdvstknprtvtvnenflslqlpsihdgdwldflslskrlvtlrglspafirfggkrt 120  
Qy 121 DFLQFQNLRNPAKSRGGPGPDYLYKNYEDDIVRSDVALDKQKCKIAQHPDVMLELQREK 180  
Db 121 dflqfqnlnrpakargpgpdylyknyeddivrsdvaldkqkckiaqhpdmlelqrek 180  
Qy 181 AAQMHLVLLKEQFSNTYSNLIIT----- 203  
Db 181 aaqmhlvllkeqfsntysnliitarsldklynfadcsghlilfalnlnrrpnnswnss 240  
Qy 204 -----EPNNYRTMHGRAVNGSOLGKDYIQLKSLLOPIRIYSRAS 242  
Db 241 alsllkysaskyniswelgnepnnyrtmhgravnsglqkdyiqkllqpiriysras 300  
Qy 243 LYGNIGRPRKNVIALLDGFMKVGSTVDVYTHQHCYIDGRVVKVMDFLKRLDFTLSQ 302  
Db 301 lygnigrprknvialldgfmkvagstvdtvthqcyidgrvvkvmfklrldtldsq 360  
Qy 303 IRKIQKVNTYTPGKKIWLQGVVTSAGTNNLSDSYAAGFLWNLGMLANQGDVIR 362  
Db 361 irkigkvntytpgkklwlegvvttsagtnnlsdsyaagflwntlgmlanqgidvir 420  
Qy 363 HSFFDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVGAGLQKPRGVRIDKLRI 422  
Db 421 hsfddhgyhnlvdqnfnpdpdywlsllkrligpkvlavhvaglgkprgvrirdklri 480  
Qy 423 YAHCTNNHHNHYVRSITFLIINLHRSRKKIKLAGTLRDLKHVHLYLQPYGQGLKSKSV 482  
Db 481 yahctnnhnhnyvrsitflinlhrsrrkkiklagtlrdklvhqylqpygqglksksv 540  
Qy 483 QLNGQPLVMVDGTLPELKPRLRAGRTLVIPTVTMGFFVVKVNNALACRYR 534  
Db 541 qlngqplvmvddgtlpeikprlragrtlviptvtmgffvkvnnalacryr 592

RESULT 3

AAB85215  
ID AAB85215 standard; Protein; 592 AA.  
XX  
AC AAB85215;  
XX  
DT 07-SEP-2001 (first entry)  
XX  
DE Heparanase-like protein Hpa2 splice variant #1.  
XX  
KW Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;  
KW cytosolic; neuroprotective; cerebroprotective; immunosuppressive;  
KW antipsoriatic; nontropic; antiinflammatory; antiarthritic; antiasthmatic;  
KW antidiabetic; antiarteriosclerotic; vulnerary.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 237 /label= unknown  
FT /note= "encoded by ANC"  
XX  
PN WO200146392-A2.  
XX  
PD 28-JUN-2001.  
XX  
PF 21-DEC-2000; 2000WO-GB04963.  
XX  
PR 22-DEC-1999; 99GB-0030392.  
PR 07-APR-2000; 2000GB-0008713.  
XX  
PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
XX  
PI McKenzie EA, Stamps AC, Terrett JA, Tyson KL;  
XX WPI; 2001-418056/44.  
DR N-PSDB; AAB22671.  
XX

PT Novel homologs of heparanase, present in three splice variants, useful  
PT for identifying agents that modulate heparanase, useful in the  
PT treatment and/or prophylaxis of abnormal levels of heparanase -  
XX Claim 1; Fig 1; 97pp; English.  
XX  
CC The invention provides a homologue to heparanase which is present in  
CC three splice variants. The heparanase homologue polypeptide is useful in  
CC the treatment of a human or non-human animal or for use in diagnosis.  
CC Vectors comprising the heparanase homologue polynucleotides are useful in  
CC the transformation or transfection of a prokaryotic or eukaryotic host.  
CC The modulators of the polypeptide are useful in the manufacture of a  
CC medicament for the treatment and/or prophylaxis of a condition/disease  
CC associated with abnormal levels of the heparanase homologue, including  
CC cancer, central nervous system (CNS) and neurodegenerative diseases,  
CC cardiovascular diseases such as stenosis following angioplasty and  
CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,  
CC allografts, inflammatory diseases, arthritis, vascular restenosis,  
CC tumour growth and progression, asthma, Alzheimer's disease, diabetic  
CC retinopathy, wound healing and inflammation. The polypeptide is also  
CC useful in diagnosis and research. The present sequence represents the  
CC amino acid sequence of the largest splice variant of the heparanase-like  
CC protein Hpa2 of the invention.  
XX  
SQ Sequence 592 AA;  
  
Query Match 98.6%; Score 2746; DB 22: Length 592;  
Best Local Similarity 90.2%; Pred No. 9,1e-270;  
Matches 534; Conservative 0; Mismatches 0; Indels 58; Gaps 1;  
  
Qy 1 MRVLCAFPEAMPSSNRPPACIAPGALYLAALLHLSSLSSQAGDRRLPVDRAAGLKEKTL 60  
Db 1 mrvlcafpeampssnrppacilapgalyallhlhslssqagdrprlpvdraaglkektl 60  
  
Qy 61 ILLDVSTKNPRTVTNENFLSLQLDPSIHDGWLDFLSSKRLVTLARGLSPAFLRGGKRT 120  
Db 61 illdvstknprtvtvnenflslqlpsihdgdwldflslskrlvtlrglspafirfggkrt 120  
  
Qy 121 DFLQFQNLRNPAKSRGGPGPDYLYKNYEDDIVRSDVALDKQKCKIAQHPDVMLELQREK 180  
Db 121 dflqfqnlnrpakargpgpdylyknyeddivrsdvaldkqkckiaqhpdmlelqrek 180  
  
Qy 181 AAQMHLVLLKEQFSNTYSNLIIT----- 203  
Db 181 aaqmhlvllkeqfsntysnliitarsldklynfadcsghlilfalnlnrrpnnswnss 240  
  
Qy 204 -----EPNNYRTMHGRAVNGSOLGKDYIQLKSLLOPIRIYSRAS 242  
Db 241 alsllkysaskyniswelgnepnnyrtmhgravnsglqkdyiqkllqpiriysras 300  
  
Qy 243 LYGNIGRPRKNVIALLDGFMKVGSTVDVYTHQHCYIDGRVVKVMDFLKRLDFTLSQ 302  
Db 301 lygnigrprknvialldgfmkvagstvdtvthqcyidgrvvkvmfklrldtldsq 360  
  
Qy 303 IRKIQKVNTYTPGKKIWLQGVVTSAGTNNLSDSYAAGFLWNLGMLANQGDVIR 362  
Db 361 irkigkvntytpgkklwlegvvttsagtnnlsdsyaagflwntlgmlanqgidvir 420  
  
Qy 363 HSFFDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVGAGLQKPRGVRIDKLRI 422  
Db 421 hsfddhgyhnlvdqnfnpdpdywlsllkrligpkvlavhvaglgkprgvrirdklri 480  
  
Qy 423 YAHCTNNHHNHYVRSITFLIINLHRSRKKIKLAGTLRDLKHVHLYLQPYGQGLKSKSV 482  
Db 481 yahctnnhnhnyvrsitflinlhrsrrkkiklagtlrdklvhqylqpygqglksksv 540  
  
Qy 483 QLNGQPLVMVDGTLPELKPRLRAGRTLVIPTVTMGFFVVKVNNALACRYR 534  
Db 541 qlngqplvmvddgtlpeikprlragrtlviptvtmgffvkvnnalacryr 592





CC associated with neoangiogenesis and metastasis related to malignant  
CC tumour formation. Heparanase-2 polynucleotides and proteins are useful as  
CC vaccines for inducing an immunological response against autoimmune  
CC disorders, blood coagulation disorders, cancer, diabetes, ischaemia,  
CC sepsis, stroke, cardiovascular diseases, or thrombosis, as well as in  
CC diagnosing (the susceptibility of a subject to) these diseases.  
CC Heparanase-2 fragments may be used as immunogens to produce antibodies  
CC immunospecific to the polypeptides, and to identify membrane bound  
CC soluble receptors, agonists or antagonists that compete with the binding  
CC of the polypeptide to the receptors. An antibody specific for  
CC heparanase-2 can be used in the diagnosis of the above diseases and in  
CC isolating or identifying clones expressing heparanase-2. The present  
CC sequence represents heparanase-2. Three regions of heparanase-2 with high  
CC immunogenicity (immunogenic epitopes) can be used to raise antibodies  
CC against heparanase-2.  
XX  
SQ Sequence 592 AA;

Query Match 98.2%; Score 2736; DB 22; Length 592;  
Best Local Similarity 89.9%; Pred. No. 9.4e-269;  
Matches 532; Conservative 1; Mismatches 1; Indels 58; Gaps 1;

Qy 1 MRVLCAPPEAMPSSSRPACIACALYALLLHLSLSSQACDRPLPVDRAAGLKEKTL 60  
Db 1 mrvlcafpampsssrpaciapagalyalallhllslssqagdrplpvdraaglikekl 60  
Qy 61 ILDDVSTKNPRTVNEFSLQLDPSIITHDGLDLSKRLVTLARGSPAFLRFGGKRT 120  
Db 61 illdvstknprtvenfslqldpsihdglwdfilskrlvtlargspafirfggkrt 120  
Qy 121 DFLQFONLRNPAKSGGPGDYLYKNYEDDIVRSDVALDKQKCKIAQHPDVMLELQREK 180  
Db 121 dflqfnlrnpaksggpgdylyknyeddivrsdvaldkqkckiaqhpdmvlvlgrek 180  
Qy 181 AAQHMLVLLKQFSNTYNNLIIT----- 203  
Db 181 aaqhmvlvllkqfsntynllitarsldklynfadcsglhlfnalnrrpnnswnss 240  
Qy 204 -----EPNNYRTMHGRAVNGSOLGKDYIOLKSLLOPIRYSRAS 242  
Db 241 alsllkysasknyslswelgnepnnyrtmhgravgsglqkdyqlklsllqpirysras 300  
Qy 243 LYGNPIGRPRKNVIALDGFMKVAGSTVDVAVTWHQCYIDGRVVKVMDFLKRLTLSDQ 302  
Db 301 lygnpiqrprknvialldgfmkvagstvdavtwhqcyidgrvvkvmfdlklrldtldsdq 360  
Qy 303 IRKIQKVVNTYTPGKKIWEGVVTTSGCTNNLSDSYAGFLWNTLGLANQGDIVVIR 362  
Db 361 irkikqvntytpgkklwlegvvttsaggttnlssdsyagflwntlgmlanqgdvivr 420  
Qy 363 HSFFDHGYNHLVDQNFPLPDYVWLSLLYKRLTGPKVAVHAGLQKPRGVRIRDKLRI 422  
Db 421 hsfdfhgyhnlvdqnfplpdyslllykrltgpkvavhaglvagrkprgvrirdklri 480  
Qy 423 YAHCTNNHNNHYVRSITFLINLHRSRKKIKLACTRLDKLVHQYLLQPYGQGLKRSV 482  
Db 481 yahctnnhnnhyvrsitflinlhrsrrkkiklactrldklvhqyllqpygqglkrsv 540  
Qy 483 QLNGQPLVMVDGTLPELKPRLRAGRILVIPPVTMGFFVKNVNALACRYR 534  
Db 541 qlngqplvmvdgtlpeikprlragrtlivppvtmgffvknvnalacryr 592

RESULT 6  
RAB84664  
SD AAB84664 standard; Protein; 492 AA.

XX AAB84664;  
XX  
DT 05-SEP-2001 (first entry)  
XX

DE Amino acid sequence of human heparanase-like polypeptide.

XX Human; heparanase-like polypeptide; gene therapy; cancer; angiogenesis;  
KW trauma; autoimmune disease; skin disease; cardiovascular disease;  
KW nervous system disease; inflammation; arthritis; genitalia;  
XX male fertility; erectile dysfunction.  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 407  
FT /note= "unspecified residue encoded by KCA"  
XX WO200148161-A2.  
XX 05-JUL-2001.  
XX  
XX 18-DEC-2000; 2000WO-EP12909.  
XX  
PR 23-DEC-1999; 99EP-0125831.  
XX  
PA (SCHD ) SCHERING AG.  
XX  
PI Siemeister G, Weiss B;  
XX  
DR WPI; 2001-418259/44.  
DR N-PSDB; AAH28347.  
XX  
XX Human Heparanase-like polynucleotide encoding polypeptides useful for  
PT modulating expression of the polypeptide and for treating cancer, -  
PT cancer metastasis, aberrant angiogenesis by gene therapy technique -  
XX  
PS Claim 9; Page 30; 30pp; English.  
XX  
XX The present sequence represents a human heparanase-like polypeptide.  
CC Heparanase-like polynucleotides are useful as a source of probes,  
CC primers and antisense molecules, and in gene therapy. Heparanase-like  
CC polynucleotides and polypeptides are useful for treating several  
CC disorders e.g., cancer, cancer metastasis. The oligonucleotides are  
CC also useful as diagnostic markers for the diagnosis of disorder such  
CC as cancer, cancer metastasis and aberrant angiogenesis. They may also  
CC act as diagnostic markers for diagnosis of disorder such as cancer,  
CC cancer metastasis and aberrant angiogenesis. The heparanase polypeptides  
CC and polynucleotides are also useful for treating trauma, autoimmune  
CC diseases, skin diseases, cardiovascular diseases, nervous system  
CC diseases, and inflammation including arthritis. Since the polynucleotide  
CC is preferentially expressed in male genitalia, modulation of its  
CC expression and/or activity may be used for medical intervention in male  
CC genitalia function that is male fertility control, erectile dysfunction.  
XX  
SQ Sequence 492 AA;

Query Match 92.1%; Score 2564; DB 22; Length 492;  
Best Local Similarity 99.6%; Pred. No. 2.1e-251;  
Matches 490; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 43 DRRPLPVDRAAGLKEKTLILLDVSTKNPRTVNEFSLQLDPSIITHDGLDLSKRLV 102  
Db 1 drrplpvdraaglikektlilldvstknprtvenfslqlpsihdglwdfilskrlv 60  
Qy 103 TLARGSPAFLRFGGKRTDFLQFONLRNPAKSGGPGDYLYKNYEDDIVRSDVALDKQK 162  
Db 61 tlargspafirfggkrtdfqlqfnlrnpaksggpgdylyknyeddivrsdvaldkqk 120  
Qy 163 GCKIAQHPDVMLELOBEKAAQHMLVLLKQFSNTYNNLIITEPNNYRTMHGRAVNGSOLG 222  
Db 121 gckiaqhpdmvllelgrekaagmhlvllkqfsntynllitpepnnyrtmhgravgsglq 180  
Qy 223 KDYIOLKSLLOPIRYSRASLYGPNIGRPRKNVIALDGFMKVAGSTVDVAVTWHQCYIDG 282  
Db 181 kdyiolkssllopiriyrslaslygpnigrprknvialldgfmkvagstvdavtwhqcyidg 240  
Qy 283 RVVKVMDFLKRLDITLSDQIRKIQKVVNTYTPGKKIWEGVVTTSGCTNNLSDSYAG 342

Db 241 rvkvmdfiktldltsqirkiqkvntypgkklwvvttsagtnnlsdsyaag 300  
QY 343 FWLWNTLGLMANQGDVIRHSHFFDHGYNHLYDQNFPLPDYWLSLYKRLIGPKVLAVH 402  
Db 301 flwIntlgmlanqgidvrlrshffdhgynhlyvdgnfnplpdywlslykrligpkvlnavh 360  
QY 403 VAGLQKRPGRVIRDKLRIYAHCTNHHNNHNYVRSITLFIINLHRSRKKIKLAGTLRDK 462  
Db 361 vaelqtkprrgvirdklriyahctnhhnnhnyvrsitlfiinlhrxkklagtlrdrk 420  
QY 463 LVHQYLLQPGYGBGLKSKSVQVNGQPLVMVDDGTLPDLKPRPLRAGRRLVIPPVTMGFFV 522  
Db 421 lvhqyllqpvggellkskvqngqplvmvddgtlpelkprlragrtlivppvtmgfyy 480  
QY 523 VKNVNALACRYR 534  
Db 481 vknvnalacryr 492  
RESULT 7  
ID AAY97633 standard; Protein; 538 AA.  
XX AC AAY97633;  
XX DT 20-APR-2001 (first entry)  
XX Human heparanase, hnp1 pn9 form, protein sequence.  
DE DE  
XX Heparanase; hnp1; wound healing; angiogenesis; restenosis; Sclera;  
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;  
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;  
KW gene therapy; human.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
FT Misc-difference 305  
FT /note= "encoded by GAC"  
XX PN WO200100643-A2.  
XX PD 04-JAN-2001.  
XX PF 19-JUN-2000; 2000WO-IL00358.  
XX PR 25-JUN-1999; 99US-0140801.  
XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
XX PI Pecker I, Michal I, Itzhaki H;  
XX DR WPI; 2001-137930/14.  
XX DR N-PSDB; AAA91098.  
XX New polynucleotides and polypeptides that are distantly homologous to  
PT heparanase, useful in wound healing, as well as in gene therapy  
PT protocols for angiogenesis, restenosis, atherosclerosis, or  
PT inflammation -  
XX Claim 10; Page 61-62; 67pp; English.  
XX This sequence represents a heparanase of the invention.  
CC The heparanase DNA and protein sequences are useful in wound healing,  
CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary  
CC diseases, neurodegenerative diseases (such as Sclera, Alzheimer's  
CC disease, and Creutzfeldt-Jakob disease) or viral infections. The  
CC heparanase coding sequence is particularly useful in gene therapy.  
XX Sequence 538 AA;  
SQ

Query Match 89.2%; Score 2485; DB 22; Length 538;  
Best Local Similarity 91.1%; Pred. No. 2.7e-243;  
Matches 493; Conservative 2; Mismatches 36; Indels 10; Gaps 4;  
QY 1 MRVLCAPPEAMPSSNSRPPACLPAGALYLALLHLSSQAGRRRLPVDRAAGLKEKTL 60  
Db 1 mrvlcappeampssnsrppaclpagallyallhlslssqagdrplpvdraaglkectl 60  
QY 61 ILLDVSTKNPVRTVNEFNLSQLQDPSIIHGWLDLSSKRLVTLARGLSAPFLRFGGKRT 120  
Db 61 illdvstknprtvtnefnlsqldpsiihgwldfllsskrlvltlargspafirfggkrt 120  
QY 121 DFLQFQNLNRPKSRGGPGPDYILKNYEDDIIVRSDVALDKQCKIAQHPDVMLELOREK 180  
Db 121 dflqfqnlnrpaksgpgpdyilknyed--arsldklynfadcsglhlfaalnrlrn 177  
QY 181 A-----AQMHLVLLKEQFSNTY--SNLILTEPNRYRTMHGRAVNGSOLGKDYIOLKSLQ 233  
Db 178 pnnswsssalilkyasaskkyniswelgnepnnrctmhgravnsgqlgkdyilqlksllq 237  
QY 234 PIRIYSRASLYGPNIGRPRKNVIALLDGFMKVAGSTVDVATWQHCHYIDGRVYKVMDFLKT 293  
Db 238 piryisraslygpnigrprknvialldgfmkvagstvdavtwqbcyldgrvkvkvmdfikt 297  
QY 294 RLDDTLSDQIRKIQKVNTYTPGKKIWLGVVTSAGTNLSDSYAAGFLWNLTLGMLA 353  
Db 298 rlldtlsaqrkikqvntytpgkklwlvgttsagtnnlsdsyaagflwntlgmla 357  
QY 354 NOGIDVIRHSHFFDHGYNHLYDQNFPLPDYWLSLYKRLIGPKVLAVHAGLQKRPQG 413  
Db 358 ngidvvrhshffdhgynhlyvdqnfplpdywlslykrligpkvlavhaglrqkprpg 417  
QY 414 RVIRDKLRIYAHCTNHHNNHNYVRSITLFIINLHRSRKKIKLAGTLRDKLVHQYLLQPGY 473  
Db 418 rvirdklriyahctnhhnnhnyvrsitlfiinlhrsrkkiklagtlrklvhqyllqpgy 477  
QY 474 QEGLSKSVQVNGQPLVMVDDGTLPDLKPRPLRAGRRLVIPPVTMGFFVKNVNALACRY 533  
Db 478 qeglkskvqngqplvmvddgtlpelkprlragrtlivppvtmgffvknvnalacry 537  
QY 534 R 534  
Db 538 r 538  
RESULT 8  
AAU07418  
ID AAU07418 standard; Protein; 480 AA.  
XX AC AAU07418;  
XX DT 18-DEC-2001 (first entry)  
XX Novel human extracellular matrix (ECM) protein #1.  
DE DE  
XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
KW antiproliferative; cardiant; vasotropic; cerebroprotective; nootropic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;  
KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;  
KW wound healing; food additive.  
XX OS Homo sapiens.  
XX WO200179253-A1.  
XX 25-OCT-2001.  
XX 11-APR-2001; 2001WO-US11643.  
XX 18-APR-2000; 2000US-198123P.  
PR

XX (HUMA-) HUMAN GENOME SCI INC.  
PI Fiscella M, Shi Y, Ebner R, Ruben SM;  
XX WPI; 2001-611720/70.  
DR N-PSDB; AAS13843.  
XX New nucleic acids encoding extracellular matrix polypeptides, for  
PT diagnosing, treating, preventing or ameliorating human disorders and  
PT disease, such as, autoimmune, hyperproliferative or cardiovascular  
PT disorders -  
XX Claim 1; Page 292-293; 308pp; English.  
XX The invention relates to novel isolated polynucleotides (I) encoding  
CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by  
CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.  
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility to  
CC a pathological condition. The antibodies to the polypeptides can also be  
CC used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities. The present sequence represents the amino acid sequence  
CC of novel human extracellular matrix (ECM) protein #1.  
XX Sequence 480 AA;  
SQ

Query Match 89.0%; Score 2478; DB 22; Length 480;  
Best Local Similarity 89.9%; Pred. No. 1.1e-242;  
Matches 480; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 MRVLCAPPEAMPSSNRPPACIAPGALYLALLHLHLSSQAGDRPLPVDRAAGLKEKTL 60  
Db 1 mrvlcafpampssnrppaciapgalylallhlhlssqagdrplpvdraaglkectl 60  
QY 61 ILLDVSTKPNRTVNFSLQLDPSIIHDGWLDFLSSRLVTLARGLSPAFIRFGKRT 120  
Db 61 illdvstknprvtvnfnslqldpsiihdgwlfdflssrklvltlrgslpafirfggkrt 120  
QY 121 DFLQFONLRNPAKSRGPGPDYLYKNYEDIVRSVDALDKQCKIAQHPDVMLELQREK 180  
Db 121 dflqfqnlnrpaksrpgpgpdylyknyedivrsvdaldkqckiaqhpdvmlelqrek 180  
QY 181 AQAMHLVLLKEQFSNTYSNLIITEPNRYRTMHGRAVNSQLGRDYIQLKSLLPRIYSR 240  
Db 150 -----epnyrtmhgravnsqglgkdyqlksllqpriysr 186  
QY 241 ASLYGNIGRPKNVTALLDGMKAVGSTVDATWQHCVIDGRVVKVWDFLKRLLDTLS 300  
Db 187 aslygnigrpnkvnvtallldgmkavgstvdavtwqhcvidgrvrvkvmfdlkrlltdtls 246  
QY 301 DOIIRKIQKVVNTYTPGKKTWLEGVWTTTSAGGTTNNLSDSVAAGFLWLTNGLMANOGIDVV 360  
Db 247 dqirkiqkvvntytpgkktwlegvwttsagggttnnlsdsyaaagflwltnglmanogidvv 306  
QY 361 IRHSFFDHGYNHLVDQNFPLPDYWLSSLYKRLIGPKVLAVHVAGLQKRPGRVIRDKL 420  
Db 307 irhsffdhgynhlvdqnfplpdywlslykrligpkvlavhvaglqkrpgrvirdkl 366

QY 421 RIYAHCTNHHNHNVRGSTITLFIINLHRSRKKIKLAGTLRDLKLVHQLQYLLQPYGQEGLSK 480  
Db 367 riyahctnhnhnvrgstitlfiinlhrskkiklagtlrdklvghyqlqpygqeglsk 426  
QY 481 SVOLNGOPLVMVDGTLPELKPRLRAGRTLVIPTMGFFVVKNNALACRYR 534  
Db 427 svqlngqplvmvdgtlpeklprlragrtlviptvmgffvkvnnalacryr 480

RESULT 9  
ID AAB85217  
XX AAB85217 standard; Protein; 480 AA.  
AC AAB85217;  
XX  
DT 07-SEP-2001 (first entry)  
XX  
DE Heparanase-like protein Hpa2 splice variant #3.  
XX  
KW Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;  
KW cytosolic; neuroprotective; cerebroprotective; immunosuppressive;  
KW antipsoriatic; neutrotropic; antiinflammatory; antiarthritic; antiasthmatic;  
KW antidiabetic; antiarteriosclerotic; vulnerary.  
XX  
OS Homo sapiens.  
XX  
PN WO200146392-A2.  
XX  
PD 28-JUN-2001.  
XX  
PF 21-DEC-2000; 2000WO-GB04963.  
XX  
PR 22-DEC-1999; 99GB-0030392.  
PR 07-APR-2000; 2000GB-0008713.  
XX  
PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
XX  
PI McKenzie EA, Stamps AC, Terrett JA, Tyson KL;  
XX WPI; 2001-418056/44.  
DR N-PSDB; AAH22673.  
XX  
PT Novel homologs of heparanase, present in three splice variants, useful  
PT for identifying agents that modulate heparanase, useful in the  
PT treatment and/or prophylaxis of abnormal levels of heparanase -  
XX  
PS Claim 1; Fig 3; 97pp; English.  
XX  
CC The invention provides a homologue to heparanase which is present in  
CC three splice variants. The heparanase homologue polypeptide is useful in  
CC the treatment of a human or non-human animal or for use in diagnosis.  
CC Vectors comprising the heparanase homologue polynucleotides are useful in  
CC the transformation or transfection of a prokaryotic or eukaryotic host.  
CC The modulators of the polypeptide are useful in the manufacture of a  
CC medicament for the treatment and/or prophylaxis of a condition/disease  
CC associated with abnormal levels of the heparanase homologue, including  
CC cancer, central nervous system (CNS) and neurodegenerative diseases,  
CC cardiovascular diseases such as restenosis following angioplasty and  
CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,  
CC allografts, inflammatory diseases, arthritis, vascular restenosis,  
CC tumour growth and progression, asthma, Alzheimer's disease, diabetic  
CC retinopathy, wound healing and inflammation. The polypeptide is also  
CC useful in diagnosis and research. The present sequence represents the  
CC amino acid sequence of the smallest splice variant of the heparanase-  
CC like protein Hpa2 of the invention.  
XX  
SQ Sequence 480 AA;

Query Match 89.0%; Score 2478; DB 22; Length 480;  
Best Local Similarity 89.9%; Pred. No. 1.1e-242;  
Matches 480; Conservative 0; Mismatches 0; Indels 54; Gaps 1;



OS Homo sapiens.  
PN WO200179253-A1.  
XX  
PD 25-OCT-2001.  
XX  
PF 11-APR-2001; 2001WO-US11643.  
XX  
PR 18-APR-2000; 2000US-198123P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Fiscella M, Shi Y, Ebner R, Ruben SM;  
XX  
DR WPI; 2001-611720/70.  
XX  
XX  
PT New nucleic acids encoding extracellular matrix polypeptides, for  
PT diagnosing, treating, preventing or ameliorating human disorders and  
PT disease, such as, autoimmune, hyperproliferative or cardiovascular  
PT disorders -  
XX  
PS  
FS Disclosure; Page 13-14; 308pp; English.  
XX  
XX  
CC The invention relates to novel isolated polynucleotides (I) encoding  
CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by  
CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.  
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility to  
CC a pathological condition. The antibodies to the polypeptides can also be  
CC used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities. The present sequence represents the amino acid sequence  
CC of human heparanase-like protein.  
XX  
SQ Sequence 439 AA;

Query Match 81.5%; Score 2270; DB 22; Length 439;  
Best Local Similarity 89.0%; Pred. NO. 1.4e-221;  
Matches 439; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

OY 42 GDRRLPVDRAAGLEKTKILLDVTSTKPNRVTVNENFLSLQDPSLIHDGWLDFLSKKRL 101  
Db 1 gdrpvpdraaglkextllldvstknprvtvnenflslqdpshldgwlfdflsskrl 60  
OY 102 VTLAGSLPAFLRFGGKRTDFLQFNLRNPAKSRGGPGDYLYLKNYEDDIVRSDVALDKQ 161  
Db 61 vtlagslspafirfggkrtdfllqfnlrnpaksrpggpgdylylknedy----- 108  
OY 162 GKCKTAQHPDVMLELOREKAAQMHVLLVLEKEQFSNTYSNLIETPNRYTMHGRAVNGSQL 221  
Db 109 -----epnyrttmhgravnsgsl 126  
OY 222 GKDYIQLKSLLOPIRYSRASLYGNIGPRKNVTALLDGFMKVAGSTYDVTWQHICYD 281  
Db 127 gkdyiqklsllqpirysraslygnigprknvntalldgmfvagstvdvwtqhcycid 186  
OY 282 GRVKVMDFLKRLLDTSQIRKIQKVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAA 341  
Db 187 grvkvmdfklrlltdtsdqirkikqvntytpgkkiwlegvvttsaggtnnlstdsyaa 246

OY 342 GFLWLTIGMLANQGDIVVIRHSFFDHGYNHLVDQNFNLPDYLWLSLLYKRILGPKVLAV 401  
Db 247 gflwltigmlanqgidvvirhsffdhgynhldvqnfnpdyldwlsllkyrllgpkvllav 306  
OY 402 HVAGLQKRPGRVIRDKLRIYAHCTNHHNHNHVVRGSIITLFIINLHRSRKKIKLACTLRD 461  
Db 307 hvaglgqrkprgrvirdkrlriyahctnnhnhnhyvrgsifcfinlhrsrrkkiklactlrd 366  
OY 462 KLVHGYLLQPYGQEGLSKSVQLNGQPLVMVDDGTLPKLPRLRAGRTLVIPTVTMGFF 521  
Db 367 klvhgyllqpyggelksksvqlngqplvmvddgtlpelkprlragrtlviptvtmgff 426  
OY 522 VVKNVNALACRYR 534  
Db 427 vvkvnvalacryr 439

RESULT 12  
AAM24147  
ID AAM24147 standard; Protein; 262 AA.  
XX  
AC AAM24147;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Human EST encoded protein SEQ ID NO: 1672.  
XX  
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
KW diagnostics; forensic test; gene mapping; genetic disorder;  
KW biodiversity; gene therapy; nutrition.  
XX  
OS Homo sapiens.  
XX  
PN WO200154477-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 25-JAN-2001; 2001WO-US02687.  
XX  
PR 25-JAN-2000; 2000US-0491404.  
PR 17-JUL-2000; 2000US-0617746.  
PR 03-AUG-2000; 2000US-0631451.  
PR 15-SEP-2000; 2000US-0663870.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
PI Cao Y, Drmanac RA, Zhang J, Werhman T;  
XX  
DR WPI; 2001-476164/51.  
DR N-PSDB; AAH98806.  
XX  
PT Isolated polypeptide for treatment of diseases, diagnostics, raising  
PT antibodies and research use -  
XX  
PS Claim 20; Page 1122-1123; 1275pp; English.  
XX  
CC The present invention provides the protein and coding sequences of novel  
CC proteins from a variety of organisms, including human, dog, cat, horse,  
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
CC from the organism of interest. They can be used in diagnostics,  
CC forensics, gene mapping, identification of mutations, to assess  
CC biodiversity and for nutritional purposes. The present sequence is a  
CC protein of the invention.  
XX  
SQ Sequence 262 AA;

Query Match 47.2%; Score 1314; DB 22; Length 262;  
Best Local Similarity 98.5%; Pred. NO. 9.1e-125;  
Matches 257; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	1	MRVLCAPPEAMPSSNRRPACIAPGALYIALLLHLSLSSQAGDRPLPVDRAAGLKEKTL	60	PR	14-AUG-2000;	2000US-0225268.
Db	1	mrvcalfpeampssnrpaciapgalylallhlslssqagdrplpvdraaglektl	60	PR	14-AUG-2000;	2000US-0225270.
QY	61	ILLDVSTKPNRTVNFENFLSLQDPSIHDGWLDFLSSKRLVTLARGLSPAFLRFGGKRT	120	PR	14-AUG-2000;	2000US-0225447.
Db	61	illdvstknprtvnfnflslqdpshdglwldflsskrlvtlargspaflrfggkrt	120	PR	14-AUG-2000;	2000US-0225447.
QY	121	DFLOFQNLNRPKSRGPGPDYVLKNYEDDIIVRSVDALDKQKCKIAQHPDVMLELQREK	180	PR	14-AUG-2000;	2000US-0225757.
Db	121	dflqfnltnrpaksgpgpdvylknyeddiivrsdvaldkqkckiaqhpdmlepprek	180	PR	14-AUG-2000;	2000US-0225758.
QY	181	AAQMHVLVLEQPSNTYSNLIITPEPNYRTMHGRAVNSQLGKDYIQLKSLIOPRIYSR	240	PR	14-AUG-2000;	2000US-0225759.
Db	181	aqmhlvllkeqfntysnliitpepnymtgravnsgldgkdyiqklsllqprlrys	240	PR	14-AUG-2000;	2000US-0226279.
QY	241	ASLYGPNIGRPRKNVTALLDG	261	PR	14-AUG-2000;	2000US-0226281.
Db	241	aslygpnivprknvtallidg	261	PR	14-AUG-2000;	2000US-0226281.
RESULT	13			PR	14-AUG-2000;	2000US-0226281.
AA99905				PR	14-AUG-2000;	2000US-0226281.
ID	AA99905	standard; Protein; 214 AA.		PR	14-AUG-2000;	2000US-0226281.
XX	AA99905;			PR	14-AUG-2000;	2000US-0226281.
DT	07-JAN-2002	(first entry)		PR	14-AUG-2000;	2000US-0226281.
XX	Human excretory related polypeptide SEQ ID NO 642.			PR	14-AUG-2000;	2000US-0226281.
DE	Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;			PR	14-AUG-2000;	2000US-0226281.
KW	immunosuppressive; antitumorigenic; anti-HIV; antibacterial; vulnerary;			PR	14-AUG-2000;	2000US-0226281.
KW	antiparkinsonian; antisickling; antianaemic; antiarthritis; cancer;			PR	14-AUG-2000;	2000US-0226281.
KW	antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;			PR	14-AUG-2000;	2000US-0226281.
KW	antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;			PR	14-AUG-2000;	2000US-0226281.
KW	antiparasitic; cardiant; immune disorder; cardiovascular disorder;			PR	14-AUG-2000;	2000US-0226281.
KW	neurological disease; infection; nephrotropic; gene therapy; vaccine;			PR	14-AUG-2000;	2000US-0226281.
XX	excretory system.			PR	14-AUG-2000;	2000US-0226281.
OS	Homo sapiens.			PR	14-AUG-2000;	2000US-0226281.
XX	WO200155313-A2.			PR	14-AUG-2000;	2000US-0226281.
PN	02-AUG-2001.			PR	14-AUG-2000;	2000US-0226281.
PD	17-JAN-2001;	2001WO-US01323.		PR	14-AUG-2000;	2000US-0226281.
XX	31-JAN-2000;	2000US-0179065.		PR	14-AUG-2000;	2000US-0226281.
PR	04-FEB-2000;	2000US-0180628.		PR	14-AUG-2000;	2000US-0226281.
PR	24-FEB-2000;	2000US-0184664.		PR	14-AUG-2000;	2000US-0226281.
PR	02-MAR-2000;	2000US-0186350.		PR	14-AUG-2000;	2000US-0226281.
PR	16-MAR-2000;	2000US-0189874.		PR	14-AUG-2000;	2000US-0226281.
PR	17-MAR-2000;	2000US-0190076.		PR	14-AUG-2000;	2000US-0226281.
PR	18-APR-2000;	2000US-0198123.		PR	14-AUG-2000;	2000US-0226281.
PR	19-MAY-2000;	2000US-0205515.		PR	14-AUG-2000;	2000US-0226281.
PR	07-JUN-2000;	2000US-0209457.		PR	14-AUG-2000;	2000US-0226281.
PR	28-JUN-2000;	2000US-0214986.		PR	14-AUG-2000;	2000US-0226281.
PR	30-JUN-2000;	2000US-0215135.		PR	14-AUG-2000;	2000US-0226281.
PR	07-JUL-2000;	2000US-0216647.		PR	14-AUG-2000;	2000US-0226281.
PR	07-JUL-2000;	2000US-0216880.		PR	14-AUG-2000;	2000US-0226281.
PR	11-JUL-2000;	2000US-0217487.		PR	14-AUG-2000;	2000US-0226281.
PR	11-JUL-2000;	2000US-0217496.		PR	14-AUG-2000;	2000US-0226281.
PR	14-JUL-2000;	2000US-0218290.		PR	14-AUG-2000;	2000US-0226281.
PR	26-JUL-2000;	2000US-0220963.		PR	14-AUG-2000;	2000US-0226281.
PR	26-JUL-2000;	2000US-0220964.		PR	14-AUG-2000;	2000US-0226281.
PR	14-AUG-2000;	2000US-0224518.		PR	14-AUG-2000;	2000US-0226281.
PR	14-AUG-2000;	2000US-0224519.		PR	14-AUG-2000;	2000US-0226281.
PR	14-AUG-2000;	2000US-0225213.		PR	14-AUG-2000;	2000US-0226281.
PR	14-AUG-2000;	2000US-0225214.		PR	14-AUG-2000;	2000US-0226281.
PR	14-AUG-2000;	2000US-0225266.		PR	14-AUG-2000;	2000US-0226281.
PR	14-AUG-2000;	2000US-0225267.		PR	14-AUG-2000;	2000US-0226281.

PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465569/50.  
XX N-PSDB; AA198878.  
XX  
XX Isolated nucleic acid molecule encoding excretory system antigen is  
XX used in preventing, treating or ameliorating a medical condition -  
XX  
XX Claim 11; SEQ ID NO 642; 574pp + Sequence Listing; English.  
XX  
XX The invention relates to novel excretory system related human  
XX polynucleotides (AA198567-AA199503) and the encoded proteins  
XX (AA199594-AA199913) useful for preventing, treating or ameliorating  
XX medical conditions e.g. by protein or gene therapy, especially  
XX disorders related to the excretory system. The genes are isolated  
XX from a range of human tissues disclosed in the specification. The  
XX nucleic acids, proteins, antibodies and (ant)agonists are useful in  
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone  
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
XX and parasitic infections.  
XX Note: The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 214 AA;

Query Match 39.9%; Score 1112; DB 22; Length 214;

Best Local Similarity 100.0%; Pred. No. 2.3e-104; Indels 0; Gaps 0;  
Matches 210; Conservative 0; Mismatches 0;

QY 325 VTTSAGTNNLSDSYAAGFLWLTGLMLANQIDVVIRHSFFDHGYNHLVDQNFNLPDY 384  
vtttsaggtnnlsdsyaagflwltglmlanqgldvvirhsffdhgynhlvdqnfnpdpdy 64  
Db 5 vtttsaggtnnlsdsyaagflwltglmlanqgldvvirhsffdhgynhlvdqnfnpdpdy 64  
QY 385 WLSLLYKRLIGPKVLAVHVAGLQKRPGRVIRDLKRIYAHCTNHHNNHYVSGSTLFTI 444  
wlslllykrligpkvlavhvaglorqkprgrvirldkrlriyahctnhhnnhyvrgstlftii 124  
Db 65 wlslllykrligpkvlavhvaglorqkprgrvirldkrlriyahctnhhnnhyvrgstlftii 124  
QY 445 NLHRSRKKIKLAGTLDKLVHQLQPYGOEGLKSKVOLNQPLVMVDDGTLPELKPRP 504  
nlhrrsrkkiklagtldkvlvhoqlqpygoeglkskvolnqplvmvddgtlpelekprp 504  
Db 125 nlhrrsrkkiklagtldkvlvhoqlqpygoeglkskvolnqplvmvddgtlpelekprp 504  
QY 505 LRAGRTLVIPTVMGFFVVKVNNALACRYR 534  
lragrtrlvipvtvmgffvvkvnvalacryr 214  
Db 185 lragrtrlvipvtvmgffvvkvnvalacryr 214

RESULT 14

AA043704  
ID AA043704 standard; Protein; 214 AA.

AC AA043704;

DT 24-OCT-2001 (first entry)

DE Human bladder antigen, SEQ ID NO: 98.

KW Human; bladder antigen; cytostatic; immunosuppressive; nootropic;  
KW neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;  
KW antiinflammatory; antitumor; antitubercular; antiparasitic; antibacterial;  
KW antifungal; antiparasitic; cardiant; gene therapy; cancer;  
KW immune disorder; cardiovascular disorder; wound healing; infection;  
KW neurological disease.

OS Homo sapiens.

PN WO200159064-A2.

PD 16-AUG-2001.

PF 17-JAN-2001; 2001WO-US01342.

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 11-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0246174.  
PR 01-NOV-2000; 2000US-0246471.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-514652/56.  
DR N-PSDB; AAI64065.  
XX  
PT Forty five bladder related polynucleotides, useful in the prevention,  
PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
disorders and neurological diseases -  
XX  
PS Claim 11; SEQ ID NO 98; 482pp + sequence listing; English.  
XX  
CC The invention relates to forty five novel bladder related  
CC polynucleotides. The polynucleotides and the polypeptides that they  
CC encode are useful in the diagnosis, treatment and prevention of:  
CC cancer, particularly breast and ovarian cancer, and other cancers  
CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal  
CC tract, liver, lung, or urogenital system; immune disorders such as  
CC Addison's disease, allergies, autoimmune haemolytic anaemia,  
CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple  
CC sclerosis, rheumatoid arthritis and ulcerative colitis; cardiovascular  
CC disorders such as myocardial ischaemias; wound healing; neurological  
CC diseases such as cerebral anoxia and epilepsy; and infectious diseases  
CC such as viral, bacterial, fungal and parasitic infections. Numerous  
CC examples of each type of disorder are given in the specification.  
CC The polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities. The polynucleotides are  
CC useful for chromosome identification. They are also useful as probes  
CC for diagnosing or treating a disorder related to the female reproductive  
CC system, particularly breast and/or ovary cancer. The present  
CC sequence is a novel bladder antigen provided in the invention.  
CC Note: The sequence data for this patent did not form part of the



CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 214 AA;

Query Match 39.9%; Score 1112; DB 22; Length 214;  
Best Local Similarity 100.0%; Pred. No. 2.3e-104;  
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 VTTSGAGTNNLSDSYAAGFLWNTLGLMLANGIDVIRHSFFDHCYNHLVDONENPLPDY 384  
Db 5 vtsaggnunlsdsyaagflwntlgmlangidvirsffdhgynhlvdqgnplpdy 64

QY 385 WLSLLYRLIGPKVLAVHAGLQRPGRVIRDKLRIYAHCTNHNHNYVRGSTITLFI 444  
Db 65 wlslllyrligpkvlavhvaglrqkprgrvirdklriyahctnhnhnyvrgstitlfi 124

QY 445 NLHSRRKKIKLAGTLRDKLVHGYLLQPGQGLSKSVQLNGQPLVMVDDGTLPDLKPRP 504  
Db 125 nlhsrrkkiklagtlrldklvhyllqpgqgllksksvqlngqplvmvddgtlpelkprp 184

QY 505 LRAGRTLVIPVTMGFFVVKVNNALACRYR 534  
Db 185 lragrtlvipvtmgffvkvnnalacryr 214

RESULT 15  
AAB08851  
ID AAB08851 standard; Protein; 535 AA.  
XX  
AC AAB08851;  
XX  
DT 15-JAN-2001 (first entry)  
XX  
DE A murine heparanase polypeptide.  
XX  
KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;  
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;  
KW wound healing; infection; burn; angiogenesis; restenosis;  
KW atherosclerosis; inflammation; neurodegenerative disease;  
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.  
XX  
OS Mus sp.  
XX  
PN WO200052178-A1.  
XX  
PD 08-SEP-2000.  
XX  
PF 14-FEB-2000; 2000WO-US03542.  
XX  
PR 01-MAR-1999; 99US-0258892.  
XX  
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
PA (FRIE/) FRIEDMAN M M.  
XX  
PI Pecker I, Vlodaysky I, Feinstein E;  
XX  
DR WPI: 2000-579289/54.  
DR N-PSDB: AAA75081.  
XX  
XX New polynucleotides encoding a polypeptide having heparanase activity,  
XX useful in wound healing and in gene therapy, particularly in treating  
XX tumour, inflammation, autoimmunity, neurodegenerative diseases  
XX  
PS Claim 22; Page 144-145; 152pp; English.  
XX

The present sequence represents murine protein with heparanase catalytic  
CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,  
CC particularly in treating tumour, inflammation or autoimmunity.  
CC Particularly, the polynucleotide is useful in modulating the  
CC bioavailability of heparin-binding growth factors, cellular responses

CC to heparin-binding growth factors (e.g. bFGF) and cytokines  
CC (e.g. interleukin (IL)-8), cell interaction with plasma lipoproteins,  
CC cellular susceptibility to certain viral and some bacterial and protozoa  
CC infections, or disintegration of neurodegenerative plaques. The  
CC polynucleotide is also useful in wound healing (e.g. thermal, chemical  
CC or radiation burns), and in the treatment of angiogenesis, restenosis,  
CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-  
CC Straussler Syndrome or Creutzfeldt-Jakob disease), and some viral,  
CC bacterial or protozoa infections.  
XX  
SQ Sequence 535 AA;

Query Match 33.9%; Score 944.5; DB 21; Length 535;  
Best Local Similarity 40.0%; Pred. No. 1.1e-86;  
Matches 211; Conservative 76; Mismatches 156; Indels 85; Gaps 9;

QY 63 LDVSTKNPVRVNVNENFLSLQDPSIIHD-GWLDFLSRRRLVTLARGLSPAFRFGGKRTD 121  
Db 33 ldytkrplrspsflsitidaslatdprflfigsprlralarglspayrlfggktld 92

QY 122 FLOFQNLNRPKSRGPGPDYVLKNVEDDIVRSDVALDKQCKIAQHPDVMLEQREKA 181  
Db 93 flif----dpdkeptseersykwqvnhdicrsepv-----saavlrklqvwp 137

QY 182 AQMHVLVLLKEQF-----SNTYS-----NLIL----- 202  
Db 138 fq-elllliregqkfstkystysvdmlysfakcsgidlfiglnallrtptdlrnssna 196

QY 203 -----TEPNRYTMHGRAVNGSQGLKDYIQLKSLLOPRTIYRSASL 243  
Db 197 qlldycsskgylniswelgnepnstfwkahlidqligedfvelhklqr-safqnkl 255

QY 244 YGNPIGRPRKNVIALLDGFMKVAGSTVDAVTWQHCHYIDGRVVKVMDLTKRLDLSDOI 303  
Db 256 yspdigprgktvklrsflkaggevidstwhhyylngriatkeofissdaldtflsv 315

QY 304 RKIQKVVNTYTPGKKIWLEGVVTTSGGTNNLSDSYAAGFLWNTLIGMLANGIDVIRH 363  
Db 316 qkilkvtkeitpgkkvwlgetsayggapllsntfaagfmlklgisagmgievvmrq 375

QY 364 SPFDHGYNHLVDQNFNPLPDYWLWLLYKRLIGPKVLAVHAGLQRPGRVIRDKLRIY 423  
Db 376 vffgagnyhlvdnfeipdywlsllfkklvgprvlsvrkqpd-----rsklrvy 426

QY 424 AHCTNHHNHNHYVRGSTITLFIINLHRSRKKIKLAGTLRDKLVHGYLLQPGQGLSKSVQ 483  
Db 427 lhctnvyptryqegditlyvlnhvtkhikvppplfrkpvdtlylkpsgpgdglssvsq 486

QY 484 LINGQPLVMVDDGTLPDLKPRPLRAGRTLVIPVTMGFFVVKVNNALAC 531  
Db 487 lngqilkmvdeqtlpalteklpagsalslpafsygffvirknakiac 534

Search completed: July 30, 2002, 08:13:59  
Job time: 336 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:13:59 ; Search time 141.47 Seconds  
(without alignments)  
69.092 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_42\_129

Perfect score: 448

Sequence: 1 GDRRLPVDRAAGLKEKTLI.....PAFLRFGKRTDFLQFNLR 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq\_032802.\*

- 1: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1980.DAT.\*
- 2: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1981.DAT.\*
- 3: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1982.DAT.\*
- 4: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1983.DAT.\*
- 5: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1984.DAT.\*
- 6: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1985.DAT.\*
- 7: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1986.DAT.\*
- 8: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1987.DAT.\*
- 9: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1988.DAT.\*
- 10: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1989.DAT.\*
- 11: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1990.DAT.\*
- 12: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1991.DAT.\*
- 13: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1992.DAT.\*
- 14: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1993.DAT.\*
- 15: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1994.DAT.\*
- 16: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1995.DAT.\*
- 17: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1996.DAT.\*
- 18: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1997.DAT.\*
- 19: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1998.DAT.\*
- 20: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1999.DAT.\*
- 21: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA2000.DAT.\*
- 22: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	448	100.0	262	22	AAW24147 Human EST encoded
2	448	100.0	439	22	AAU07423 Human heparanase-1
3	448	100.0	480	22	AAU07418 Novel human extrac
4	448	100.0	480	22	AAW85217 Heparanase-like pr
5	448	100.0	480	22	AAU97634 Human heparanase,
6	448	100.0	534	22	AAW85216 Heparanase-like pr
7	448	100.0	538	22	AAU97633 Human heparanase,
8	448	100.0	592	22	AAU07424 Human heparanase-1
9	448	100.0	592	22	AAW85215 Heparanase-like pr
10	448	100.0	592	22	AAW81062 Human Heparanase-2
11	448	100.0	592	22	AAU97632 Human heparanase,

12	442	98.7	492	22	AAW84664 Amino acid sequenc
13	166.5	37.2	535	21	AAW08851 A murine heparanas
14	150.5	33.6	532	20	AAU17083 Seg ID No: 15 of W
15	150.5	33.6	543	20	Human heparanase e
16	150.5	33.6	543	20	AAU02345 A human heparanase
17	150.5	33.6	543	21	AAW08849 Amino acid sequenc
18	150.5	33.6	543	21	AAU57590 Human heparanase. p
19	150.5	33.6	543	21	AAU52990 Human heparanase i
20	150.5	33.6	543	22	AAW86206 Human heparanase o
21	150.5	33.6	543	22	AAW88361 Human membrane or
22	150.5	33.6	543	22	AAU97635 Human heparanase p
23	150.5	33.6	588	20	AAU30124 A human protein wi
24	150.5	33.6	592	20	AAU02346 A human heparanase
25	150.5	33.6	592	21	AAW08850 Amino acid sequenc
26	149.5	33.4	530	20	AAU34173 Human pre-prohepar
27	93	20.8	32	20	AAU34175 Human pre-prohepar
28	84.5	18.9	488	22	AAW31469 Amino acid sequenc
29	82.5	18.4	488	22	AAW31472 Amino acid sequenc
30	80.5	18.0	488	22	AAW31470 Amino acid sequenc
31	76.5	17.1	488	22	AAW31471 Amino acid sequenc
32	72.5	16.2	280	21	AAW11311 Arabidopsis thalia
33	72.5	16.2	280	21	AAW45380 Arabidopsis thalia
34	72.5	16.2	302	21	AAW11310 Arabidopsis thalia
35	72.5	16.2	302	21	AAW45379 Arabidopsis thalia
36	72.5	16.2	304	21	AAW11309 Arabidopsis thalia
37	72.5	16.2	304	21	AAW45378 Arabidopsis thalia
38	68.5	15.3	211	21	AAW05952 Protein deduced fr
39	64	14.3	349	21	AAW17097 Arabidopsis thalia
40	64	14.3	369	21	AAW24233 Arabidopsis thalia
41	64	14.3	369	21	AAW47069 Arabidopsis thalia
42	64	14.3	405	21	AAW47028 Arabidopsis thalia
43	64	14.3	2273	17	AAW98811 Erysiphe graminis
44	63.5	14.2	935	22	ABW69219 Drosophila melanog
45	63.5	14.2	1306	19	AAW68155 Human angiotensin

ALIGNMENTS

RESULT	1
AAW24147	
ID	AAW24147 standard; Protein; 262 AA.
AC	AAW24147;
XX	
DT	12-OCT-2001 (first entry)
XX	
DE	Human EST encoded protein SEQ ID NO: 1672.
XX	
KW	Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW	Tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW	diagnostics; forensic test; gene mapping; genetic disorder;
KW	biodiversity; gene therapy; nutrition.
XX	
OS	Homo sapiens.
XX	
PN	WO200154477-A2.
PD	02-AUG-2001.
XX	
PF	25-JAN-2001; 2001WO-US02687.
XX	
PR	25-JAN-2000; 2000US-0491404.
PR	17-JUL-2000; 2000US-0617746.
PR	03-AUG-2000; 2000US-0631451.
PR	15-SEP-2000; 2000US-0663870.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI	Cao Y, Drmanac RA, Zhang J, Werhman T;
XX	
DR	WPI; 2001-476164/51.

DR N-PSDB; AAH98806.

XX Isolated polypeptide for treatment of diseases, diagnostics, raising

PT antibodies and research use -

XX Claim 20; Page 1122-1123; 1275pp; English.

XX The present invention provides the protein and coding sequences of novel

CC proteins from a variety of organisms, including human, dog, cat, horse,

CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea

CC urchin and tomato. These were derived from expressed sequence tags (ESTs)

CC from the organism of interest. They can be used in diagnostics,

CC forensics, gene mapping, identification of mutations, to assess

CC biodiversity and for nutritional purposes. The present sequence is a

CC protein of the invention.

SQ Sequence 262 AA;

Query Match 100.0%; Score 448; DB 22; Length 262;

Best Local Similarity 100.0%; Pred. No. 2.3e-48;

Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIIHDGWLDFLSSKRL 60

DB 42 GDRRLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIIHDGWLDFLSSKRL 101

QY 61 VTLAGLSPAFLRFGGKRTDFLQFNLR 88

DB 102 VTLAGLSPAFLRFGGKRTDFLQFNLR 129

RESULT 2

AAU07423

ID AAU07423 standard; Protein: 439 AA.

XX AC AAU07423;

XX 18-DEC-2001 (first entry)

XX Human heparanase-like protein.

XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

KW antiproliferative; cardiant; vasotropic; cerebroprotective; nootropic;

KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;

KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;

KW hyperproliferative disorder; neoplasm; cardiovascular disorder;

KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;

KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;

KW wound healing; food additive; heparanase.

XX Homo sapiens.

XX OS

XX PN WO200179253-A1.

XX PD 25-OCT-2001.

XX 11-APR-2001; 2001WO-US11643.

XX 18-APR-2000; 2000US-198123P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Fiscella M, Shi Y, Ebner R, Ruben SM;

XX WPI; 2001-611720/70.

XX New nucleic acids encoding extracellular matrix polypeptides, for

PT diagnosing, treating, preventing or ameliorating human disorders and

PT disease, such as, autoimmune, hyperproliferative or cardiovascular

PT disorders -

XX Disclosure; Page 13-14; 308pp; English.

XX

XX The invention relates to novel isolated polynucleotides (I) encoding

CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by

CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.

CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They

CC are also used in diagnosing a pathological condition or susceptibility to

CC a pathological condition. The antibodies to the polypeptides can also be

CC used in alleviating symptoms associated with the disorders and in

CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked

CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated

CC include autoimmune diseases e.g. rheumatoid arthritis,

CC hyperproliferative disorders e.g. cardiac arrest, cerebrovascular disorders

CC cardiovascular disorders e.g. cardiac arrest, nervous system disorders e.g.

CC e.g. cerebral ischaemia, angiogenesis, caused by bacteria, viruses and fungi

CC Alzheimer's disease, infections caused by bacteria, viruses and fungi

CC and ocular disorders e.g. corneal infection. The polypeptides can also

CC be used to aid wound healing and epithelial cell proliferation, to

CC prevent skin aging due to sunburn, to maintain organs before

CC transplantation, for supporting cell culture of primary tissues, to

CC regenerate tissues and in chemotaxis. The polypeptides can also be used

CC as a food additive or preservative to increase or decrease storage

CC capabilities. The present sequence represents the amino acid sequence

CC of human heparanase-like protein.

XX SQ Sequence 439 AA;

Query Match 100.0%; Score 448; DB 22; Length 439;

Best Local Similarity 100.0%; Pred. No. 4.4e-48;

Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIIHDGWLDFLSSKRL 60

DB 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIIHDGWLDFLSSKRL 60

QY 61 VTLAGLSPAFLRFGGKRTDFLQFNLR 88

DB 61 VTLAGLSPAFLRFGGKRTDFLQFNLR 88

RESULT 3

AAU07418

ID AAU07418 standard; Protein: 480 AA.

XX AC AAU07418;

XX 18-DEC-2001 (first entry)

XX Novel human extracellular matrix (ECM) protein #1.

XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

KW antiproliferative; cardiant; vasotropic; cerebroprotective; nootropic;

KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;

KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;

KW hyperproliferative disorder; neoplasm; cardiovascular disorder;

KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;

KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;

KW wound healing; food additive.

XX Homo sapiens.

XX OS

XX PN WO200179253-A1.

XX PD 25-OCT-2001.

XX 11-APR-2001; 2001WO-US11643.

XX 18-APR-2000; 2000US-198123P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Fiscella M, Shi Y, Ebner R, Ruben SM;

WPT: 2001-611720/70.  
N-PSDB: AAS13843.

New nucleic acids encoding extracellular matrix polypeptides, for diagnosing, treating, preventing or ameliorating human disorders and disease, such as, autoimmune, hyperproliferative or cardiovascular disorders -

Claim 1; Page 292-293; 308pp; English.

The invention relates to novel isolated polynucleotides (I) encoding extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by (I) are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. The antibodies to the polypeptides can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis,

hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. The present sequence represents the amino acid sequence of novel human extracellular matrix (ECM) protein #1.

Sequence 480 AA;

```
Query Match      100.0%; Score 448; DB 22; Length 480;
Best Local Similarity 100.0%; Pred. NO. 4.9e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	1	GDRPLPVDRAGLIKETLIILLDYSTKNPVTVNENFLSLQDPSIIHDGWLDFUSSRL	60
Db	42	gdrplpvdraagliketliilldstknpvtvnfnlsqldpslihdgwdfisskrl	101
Qy	61	VTLAGLSPAFLEFGGKRTDFLQFNLR	88
Db	102	vtlaglspafirfggkrtdflgfnlr	129

RESULT 4  
AAB85217  
ID AAB85217 standard; Protein; 480 AA.

DT 07-SEP-2001 (first entry)

DE Heparanase-like protein Hpa2 splice variant #3.

KW Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;  
 KW cytosolic; neuroprotective; cerebroprotective; immunosuppressive;  
 KW antiparasitic; nontropic; antiinflammatory; antiarthritic; antiasthmatic;  
 KW antidiabetic; antiarteriosclerotic; vulnery.

OS Homo sapiens.

PN WO200146392-A2.

PD 28-JUN-2001.

PF 21-DEC-2000; 2000WO-GB04963.

PR 22-DEC-1999; 99GB-0030392.

PR 07-APR-2000; 2000GB-0008713.

XX  
PA (OXFO-) OXFORD GLYSCSCIENCES UK LTD.

PI McKenzie EA, Stamps AC, Terrett JA, Tyson KL;

DR WPI; 2001-418056/44.

DR N-PSDB; AAH22673.

Novel homologs of heparanase, present in three splice variants, useful for identifying agents that modulate heparanase, useful in the treatment and/or prophylaxis of abnormal levels of heparanase

PS Claim 1; Fig 3; 97pp; English.

The invention provides a homologue to heparanase which is present in three splice variants. The heparanase homologue polypeptide is useful in the treatment of a human or non-human animal or for use in diagnosis. Vectors comprising the heparanase homologue polynucleotides are useful in the transformation or transfection of a prokaryotic or eukaryotic host. The modulators of the polypeptide are useful in the manufacture of a CC medicament for the treatment and/or prophylaxis of a condition/disease CC associated with abnormal levels of the heparanase homologue, including CC cancer, central nervous system (CNS) and neurodegenerative diseases, CC cardiovascular diseases such as stenosis following angioplasty and CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus, CC allografts, inflammatory diseases, arthritis, vascular stenosis, CC tumour growth and progression, asthma, Alzheimer's disease, diabetic CC retinopathy, wound healing and inflammation. The polypeptide is also CC useful in diagnosis and research. The present sequence represents the CC amino acid sequence of the smallest splice variant of the heparanase- CC like protein Hpa2 of the invention.

AA	Sequence	480 AA;
SQ		

Query Match	100.0%;	Score 448;	DB 22;	Length 480;
Best Local Similarity	100.0%;	Pred. No. 4.9e-48;		

QY 1 GDRRLPVDRRAAGLAEKTLILLDVSTKNPVRTVNENFISLQLDPSIIHDGWDLFSSKRL 60  
 |||||  
 Db 42 gdrprlpvdraaglkectllldvstknprtvenfislqldpsilhdwldflsskrl 101

QY 61 VTLARGLSPAFLRFGGKRTDFLQFQNL 88

Db 102 vtlarglspaf1rfggkrtdf1qfqnlr 129

RESULT 5

AA97634  
ID AAY97634 standard; Protein; 480 AA.

AA  
AC AAY97634;

DT 20-APR-2001 (first entry)

Human heparanase, hnhp1 pn5 form, protein sequence.

Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scurvy;  
 KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;  
 KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;  
 KW gene therapy; human.

XX  
OS  
Homo sapiens.

XX  
PN  
WO200100643-A2.

XX  
PD  
04-JAN-2001.

XX PF 19-JUN-2000; 2000WO-IL00358.

XX  
PR 25-JUN-1999; 99US-0140801.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
XX Pecker I, Michal I, Itzhaki H;  
XX WPI; 2001-137930/14.  
XX N-PSDB; AAA91099.  
XX New polynucleotides and polypeptides that are distantly homologous to  
PT heparanase, useful in wound healing, as well as in gene therapy  
PT protocols for angiogenesis, restenosis, atherosclerosis, or  
PT inflammation -  
XX  
XX Claim 10; Page 63; 67pp; English.  
XX This sequence represents a heparanase of the invention.  
CC The heparanase DNA and protein sequences are useful in wound healing,  
CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary  
CC diseases, neurodegenerative diseases (such as Scrape, Alzheimer's  
CC disease, and Creutzfeldt-Jakob disease) or viral infections. The  
CC heparanase coding sequence is particularly useful in gene therapy.  
XX  
XX Sequence 480 AA;  
SQ

Query Match 100.0%; Score 448; DB 22; Length 480;  
Best Local Similarity 100.0%; Pred. No. 4.9e-48;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHGWLDLSSKRL 60  
Db 42 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHGWLDLSSKRL 101  
QY 61 VTLAGLSPAFRLFGGKRTDFLOFQNL 88  
Db 102 VTLAGLSPAFRLFGGKRTDFLOFQNL 129

RESULT 6  
AAB85216  
ID AAB85216 standard; Protein; 534 AA.  
XX  
XX AAB85216;  
XX  
XX 07-SEP-2001 (first entry)  
XX Heparanase-like protein Hpa2 splice variant #2.  
XX  
XX Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;  
KW cytostatic; neuroprotective; cerebroprotective; immunosuppressive;  
KW antipsoriatic; nontropic; antiinflammatory; antiarthritic; antiasthmatic;  
KW antidiabetic; antiarteriosclerotic; vulnery.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200146392-A2.  
XX  
XX 28-JUN-2001.  
XX  
XX 21-DEC-2000; 2000WO-GB04963.  
XX  
XX 22-DEC-1999; 99GB-0030392.  
XX  
XX 07-APR-2000; 2000GB-0008713.  
XX  
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
XX  
XX McKenzie EA, Stamps AC, Terrett JA, Tyson KL;  
PI  
XX WPI; 2001-418056/44.  
XX  
XX N-PSDB; AAH22672.  
XX  
XX Novel homologs of heparanase, present in three splice variants, useful  
PT for identifying agents that modulate heparanase, useful in the

PT treatment and/or prophylaxis of abnormal levels of heparanase -  
XX Claim 1; Fig 2; 97pp; English.  
XX  
XX The invention provides a homologue to heparanase which is present in  
CC three splice variants. The heparanase homologue polypeptide is useful in  
CC the treatment of a human or non-human animal or for use in diagnosis.  
CC Vectors comprising the heparanase homologue polynucleotides are useful in  
CC the transfection or transfection of a prokaryotic or eukaryotic host.  
CC The modulators of the polypeptide are useful in the manufacture of a  
CC medicament for the treatment and/or prophylaxis of a condition/disease  
CC associated with abnormal levels of the heparanase homologue, including  
CC cancer, central nervous system (CNS) and neurodegenerative diseases,  
CC cardiovascular diseases such as restenosis following angioplasty and  
CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,  
CC allografts, inflammatory diseases, arthritis, vascular restenosis,  
CC tumour growth and progression, asthma, Alzheimer's disease, diabetic  
CC retinopathy, wound healing and inflammation. The polypeptide is also  
CC useful in diagnosis and research. The present sequence represents the  
CC amino acid sequence of the mid-sized splice variant of the heparanase-  
XX like protein Hpa2 of the invention.  
XX  
XX Sequence 534 AA;  
SQ

Query Match 100.0%; Score 448; DB 22; Length 534;  
Best Local Similarity 100.0%; Pred. No. 5.6e-48;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHGWLDLSSKRL 60  
Db 42 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHGWLDLSSKRL 101  
QY 61 VTLAGLSPAFRLFGGKRTDFLOFQNL 88  
Db 102 VTLAGLSPAFRLFGGKRTDFLOFQNL 129

RESULT 7  
AAY97633  
ID AAY97633 standard; Protein; 538 AA.  
XX  
XX AAY97633;  
XX  
XX 20-APR-2001 (first entry)  
XX Human heparanase, hnhp1 pn9 form, protein sequence.  
XX  
XX Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;  
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;  
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;  
KW gene therapy; human.  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FT Misc-difference 305 /note= "encoded by GAC"  
XX  
XX WO200100643-A2.  
XX  
XX 04-JAN-2001.  
XX  
XX 19-JUN-2000; 2000WO-IL00358.  
XX  
XX 25-JUN-1999; 99US-0140801.  
XX  
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
XX Pecker I, Michal I, Itzhaki H;  
XX  
XX WPI; 2001-137930/14.  
XX  
XX N-PSDB; AAA91098.  
XX

XX New polynucleotides and polypeptides that are distantly homologous to  
PT heparanase, useful in wound healing, as well as in gene therapy  
PT protocols for angiogenesis, restenosis, atherosclerosis, or  
PT inflammation -  
XX  
PS Claim 10; Page 61-62; 67pp; English.  
XX  
CC This sequence represents a heparanase of the invention.  
CC The heparanase DNA and protein sequences are useful in wound healing,  
CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary  
CC diseases, neurodegenerative diseases (such as Scrapie, Alzheimer's  
CC disease, and Creutzfeldt-Jakob disease) or viral infections. The  
CC heparanase coding sequence is particularly useful in gene therapy.  
XX  
SQ Sequence 538 AA;  
  
Query Match 100.0%; Score 448; DB 22; Length 538;  
Best Local Similarity 100.0%; Pred. No. 5.6e-48;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKPNVTVNENFLSLQDPSIIHGWLDLSSKRL 60  
Db 42 gdrplpvdraaglkettlllldvstknprtvtvnenfslqlpsihdgwidfsskrl 101  
  
QY 61 VTLAGLSPAFLRFGGKRTDFLOFQNL 88  
Db 102 vtlarglspafirfggkrtdfiqgnlr 129  
  
RESULT 8  
AAU07424  
TS AAU07424 standard; Protein; 592 AA.  
  
XX AC AAU07424;  
XX  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Human heparanase-like protein splice variant #1.  
XX  
KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
KW antiproliferative; cardiant; vasotropic; cerebroprotective; neurotropic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;  
KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;  
KW wound healing; food additive; heparanase.  
XX  
OS Homo sapiens.  
XX  
XX WO200179253-A1.  
PN  
XX  
PD 25-OCT-2001.  
XX  
XX 11-APR-2001; 2001WO-US11643.  
PF  
XX  
PR 18-APR-2000; 2000US-198123P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Fiscella M, Shi Y, Ebner R, Ruben SM;  
XX  
XX WPI; 2001-611720/70.  
DR  
XX  
DR N-PSDB; AAS13848.  
XX  
XX New nucleic acids encoding extracellular matrix polypeptides, for  
PT diagnosing, treating, preventing or ameliorating human disorders and  
PT disease, such as, autoimmune, hyperproliferative or cardiovascular  
PT disorders -  
XX  
XX Disclosure; Page 14; 308pp; English.

XX The invention relates to novel isolated polynucleotides (I) encoding  
CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by  
CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.  
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility to  
CC a pathological condition. The antibodies to the polypeptides can also be  
CC used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis.  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities. The present sequence represents the amino acid sequence  
CC of human heparanase-like protein, splice variant #1.  
XX  
SQ Sequence 592 AA;  
  
Query Match 100.0%; Score 448; DB 22; Length 592;  
Best Local Similarity 100.0%; Pred. No. 6.4e-48;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKPNVTVNENFLSLQDPSIIHGWLDLSSKRL 60  
Db 42 gdrplpvdraaglkettlllldvstknprtvtvnenfslqlpsihdgwidfsskrl 101  
  
QY 61 VTLAGLSPAFLRFGGKRTDFLOFQNL 88  
Db 102 vtlarglspafirfggkrtdfiqgnlr 129  
  
RESULT 9  
AAB85215  
ID AAB85215 standard; Protein; 592 AA.  
XX  
AC AAB85215;  
XX  
DT 07-SEP-2001 (first entry)  
XX  
DE Heparanase-like protein Hpa2 splice variant #1.  
XX  
XX Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;  
KW cytostatic; neuroprotective; cerebroprotective; immunosuppressive;  
KW antipsoriatic; neurotropic; antiinflammatory; antiarthritic; antiasthmatic;  
KW antidiabetic; antiarteriosclerotic; vulneryary.  
XX  
OS Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 237  
FT /label= unknown  
FT /note= "encoded by ANC"  
XX  
XX WO200146392-A2.  
PN  
XX  
XX 28-JUN-2001.  
XX  
XX 21-DEC-2000; 2000WO-GB04963.  
PF  
XX  
XX 22-DEC-1999; 99GB-0030392.  
PR  
XX 07-APR-2000; 2000GB-0008713.  
PR  
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
PA  
XX

XX	11-SEP-2000; 2000WO-EP08837.
PP	
XX	23-SEP-1999; 99EP-0118805.
PR	
XX	07-JUL-2000; 2000EP-0114649.
XX	(MERE ) MERCK PATENT GMBH.
XX	
XX	Duecker K. Sirrenberg C;
XX	
DR	WPI; 2001-308089/32.
DR	N-PSDB; AAF8610L.
XX	
XX	New heparanase-2 polypeptide useful in diagnosing (the susceptibility
PPT	of a subject to) and as vaccines against e.g. autoimmune disorders,
PPT	cardiovascular disease, cancer, diabetes, ischemia, sepsis, stroke, or
PT	thrombosis -
XX	
XX	Claim 1; Page 42-43; 46pp; English.
XX	
CC	This invention relates to a human heparanase-2 protein and the cDNA
CC	sequence encoding it. Heparanase-2 is a member of the endoglucuronidase
CC	family of polypeptides and it degrades heparan sulphate proteoglycans
CC	HSPGs (ubiquitous macromolecules of cell surfaces, basement membranes and
CC	the extracellular matrix). HSPGs support the vascular endothelium and
CC	stabilise the structure of the capillary wall. Heparanases may be
CC	associated with neangiogenesis and metastasis related to malignant
CC	tumour formation. Heparanase-2 polynucleotides and proteins are useful as
CC	vaccines for inducing an immunological response against autoimmune
CC	disorders, blood coagulation disorders, cancer, diabetes, ischaemia,
CC	sepsis, stroke, cardiovascular diseases, or thrombosis, as well as in
CC	diagnosing (the susceptibility of a subject to) these diseases.
CC	Heparanase-2 fragments may be used as immunogens to produce antibodies
CC	immunospecific to the polypeptides, and to identify membrane bound
CC	soluble receptors, agonists or antagonists that compete with the binding
CC	of the polypeptide to the receptors. An antibody specific for
CC	heparanase-2 can be used in the diagnosis of the above diseases and in
CC	isolating or identifying clones expressing heparanase-2. The present
CC	sequence represents heparanase-2. Three regions of heparanase-2 with high
CC	immunogenicity (immunogenic epitopes) can be used to raise antibodies
CC	against heparanase-2.
XX	
SQ	Sequence 592 AA:
	Query Match 100.0%; Score 448; DB 22; Length 592;
	Best Local Similarity 100.0%; Pred. No. 6.4e-48;
	Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY	1 GDRRLPVPDRAAGLEKTKILLIDVSTKPKVRTVNENFLSLQLDSPSIHDGWLDFLSKKRL 60
Db	42 gdrplpvdraaglkektilldvstkpkvrtvnemflslqldspilhdgwldfllskrl 101
QY	61 VTLARGLSPAFLRFSGKRFTDFLOFNLR 88
Db	102 vtlarglspafirfggkrtdflqgnlr 129
RESULT 11	
AAY97632	
ID ID	AAY97632 standard; Protein; 592 AA.
XX AC	AAY97632;
XX AC	
DT DT	20-APR-2001 (first entry)
XX XX	Human heparanase, hnhp1, protein sequence.
DE DE	
XX KW	Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
KW KW	atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW KW	neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW KW	gene therapy; human.
XX OS	Homo sapiens.









**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:16:22 ; Search time 71.68 Seconds  
(without alignments)  
117.967 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_42\_129  
Perfect score: 448  
Sequence: 1 GDRRLPLVDRAAGLKEKTLI.....PAFLRFGGKRTDFLQFQNL 88

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	448	100.0	480	JC7506	heparanase protein
2	73.5	16.4	253	AF3625	cell division inhi
3	73	16.3	481	S56299	hypothetical prote
4	72.5	16.2	334	A86397	protein T7N9.6 [im
5	68.5	15.3	694	F90609	exodeoxyribonuclea
6	67.5	15.1	293	T09171	ribosomal protein
7	67.5	15.1	293	T09170	ribosomal protein
8	65.5	14.6	1391	T20642	hypothetical prote
9	65.5	14.6	1397	E87998	protein F09C3.1 [i
10	65	14.5	2288	T30568	acetyl-CoA carboxy
11	63.5	14.2	492	S71455	catalase (EC 1.11.
12	63.5	14.2	497	T29814	hypothetical prote
13	63	14.1	1178	AC3394	DNA polymerase III
14	62.5	14.0	326	A97316	hypothetical prote
15	62.5	14.0	618	S33044	hypothetical prote
16	62.5	14.0	732	S05238	peptidyl-dipectida
17	62.5	14.0	1306	A31759	stage II sporulati
18	62	13.8	830	F83659	hypothetical prote
19	61.5	13.7	126	E71233	peptidyl-dipectida
20	61.5	13.7	299	S25782	hypothetical prote
21	61.5	13.7	299	D23696	hypothetical prote
22	61.5	13.7	302	B70631	lacX protein - Lac
23	61.5	13.7	1299	AH2090	hypothetical prote
24	61	13.6	126	B75200	two-component hybr
25	61	13.6	275	S52278	hypothetical prote
26	61	13.6	352	D97947	hypothetical prote
27	61	13.6	423	T03269	UDP-N-acetylglucos
28	61	13.6	426	B69455	uroporphyrin-III C
29	61	13.6	433	C82899	conserved hypothet
					thymidine phosphor

30	61	13.6	516	2	B64551	oligopeptide ABC t
31	61	13.6	599	2	S18735	centromere protein
32	60.5	13.5	229	2	F75078	hypothetical prote
33	60.5	13.5	280	2	H90338	hypothetical prote
34	60	13.4	240	2	G71968	hypothetical prote
35	60	13.4	240	2	B64540	conserved hypothet
36	60	13.4	337	2	B69356	hypothetical prote
37	60	13.4	442	2	H86578	YopN [imported] -
38	60	13.4	442	2	C72046	type III secretion
39	60	13.4	613	2	JC6033	mosquitocidal prot
40	60	13.4	2185	1	S60200	acetyl-CoA carboxy
41	59.5	13.3	183	2	B97784	formylmethionine d
42	59.5	13.3	363	2	S44443	RAD23 protein homo
43	59.5	13.3	395	2	F82852	conserved hypothet
44	59.5	13.3	433	2	T48118	hypothetical prote
45	59.5	13.3	735	2	148101	ADAM 6 protein pre

ALIGNMENTS

RESULT 1  
JC7506  
heparanase protein 2a - human  
C:Species: Homo sapiens (man)  
C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 01-Dec-2000  
C:Accession: JC7506  
R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hircok, M.;  
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000  
A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase fami  
A:Reference number: JC7506  
A:Accession: JC7506  
A:Molecule type: mRNA  
A:Residues: 1-480 <MCK>  
A:Cross-references: GB:AF282885  
C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and th  
therapies.  
C:Genetics:  
A:Gene: hpa2a  
A:Map position: 10q23-10q24  
C:Keywords: heparin binding; membrane bound

Query Match 100.0%; Score 448; DB 2; Length 480;  
Best Local Similarity 100.0%; Pred. No. 2.8e-42;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GDRRLPLVDRAAGLKEKTLIILDVSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSSRRL 60  
Db 42 GDRRLPLVDRAAGLKEKTLIILDVSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSSRRL 101  
Qy 61 VTLARGLSPAFLRFGGKRTDFLQFQNL 88  
Db 102 VTLARGLSPAFLRFGGKRTDFLQFQNL 129  
RESULT 2  
AF3625  
cell division inhibitor minC [imported] - Brucella melitensis (strain 16M)  
C:Species: Brucella melitensis  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 15-Feb-2002  
C:Accession: AF3625  
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanov  
; Mazur, M.; Gotsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit  
A:Reference number: AD3252; PMID:11756688  
A:Accession: AF3625  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-253 <KUR>  
A:Cross-references: GB:AF008918; PIDN:AAU54169.1; PID:gl7985135; GSPDB:GN00191  
A:Experimental source: strain 16M

```
C:Genetics:
A:Gene: BMEII0927
A:Map position: II
C:Superfamily: cell division inhibitor minC
```

	Query Match	16.4%	Score 73.5;	DB 2;	Length 253;
	Best Local Similarity	37.3%;	Pred. No. 1..2;	Mismatches 21;	Indels 9; Gaps 2;
	Matches	22;	Conservative		
QY	29	PVRTVNENFLSLQLDPSIHDGWLDFLSKKRLVTLARGLSPAFLRFGGKRTDFLQFONL	87		
	:	: ::   :	::  :		:
Dd	15	PIRLKRSFLAMVLSPDLGWLE----	RDDLARRSSGFEL-----GRPVVDMENT	64	

```
RESULT 3
```

N: Alternate names: hypothetical protein F006  
C: Species: Saccharomyces cerevisiae  
C: Date: 02-Sep-1995 sequence\_revision 19-Oct-1995 #text\_change 29-Oct-1999  
C: Accession: S562299; S62255; S63791  
R: Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasano  
A: Submitted to the EMBL Data Library, May 1995  
A: Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces ce  
A: Reference number: S56186  
A: Accession: S56299  
A: Molecule type: DNA  
A: Residues: 1-481 <MUR>

submitted to the EMBL Data Library, December 1994

A:Reference number: S62230

A:Accession: S62255

A:Molecule type: DNA

A:Residues: 1-481 <MW>

A:Cross-references: EMBL:D44597; NID:g871938; PID:d1008600; PID:g871943

R:EkT, T.; Naitou, M.; Hagiwara, H.; Ozawa, M.; Sasanuma, S.I.; Sasanuma, M.; Tsuchiya, T.

Reast 12, 145-167, 1996

A:Title: Analysis of a 36.2 kb DNA sequence including the right telomere of chromosome V

A:Reference number: S63787; MUID:96287652

A:Accession: S63791

A:Molecule type: DNA

A:Status: nucleic acid sequence not shown; translation not shown

A:Residues: 1-481 <EKI>

A:Cross-references: EMBL:D44597; NID:g871938; PIDN:BAA08010.1; PID:d1008600; PID:g871943

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1995

C:Genetics:

A:Map position: 6R

A:Note: YFR044C

Query Match	16.3%	Score 73;	DB 2;	Length 481;
Best Local Similarity	28.6%	Pred. No. 3;		
Matches 28;	Conservative 11;	Mismatches 31;	Indels 28;	Gaps 4;
QY	17	KTLLLDVSTKPNRTV	-----NENFSL-----	-----QLDPSIHG--WL-- 52
Db	334	KTVIPAKVGFKFSRTVPDMSDKLTSLVQRKCDAKFKLSNPNKCRTELHIDGAYWSD		393

**Db**

394	PFNQFTAAKATKLVLGVDPDFTRGGSIPTLTTFQD	431
-----	-------------------------------------	-----

**RESULT 4**

A86397  
protein T7N9.6 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C;Accession: A86397  
R;Theologits, A.; Ecker, J.R.; Palm, C.J.; Federspriel, N.A.; Kaul, S.; White, O.; Alonso,



ה  
ו  
ז  
ח  
ט  
י  
יא





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:31:25 ; Search time 36.14 Seconds  
(without alignments)  
94.281 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_42\_129  
Perfect score: 448  
Sequence: 1 GDRRLPVDRAAGLKEKTLI.....PAFLRFGKRTDQLQFNLR 88

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SWISSProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	73	16.3	481	1 YFL4_YEAST	P43616 saccharomyc
2	67.5	15.1	293	1 RK4_SPIOL	O49937 spinacia ol
3	63.5	14.2	491	1 CAT2_MALZE	P12365 zea mays (m
4	63	14.1	219	1 C270_MOUSE	P58468 mus musculu
5	63	14.1	987	1 SOXA_RHIME	O87386 rhizobium m
6	62.5	14.0	575	1 UL87_EBV	P25215 Epstein-bar
7	62.5	14.0	732	1 ACET_HUMAN	P22966 homo sapien
8	62.5	14.0	1067	1 HGBB_HAEIN	Q9kiv1 haemophilus
9	62.5	14.0	1306	1 ACE_HUMAN	P12821 homo sapien
10	61.5	13.7	299	1 LAXC_LACIA	P42096 lactococcus
11	61.5	13.7	299	1 LAXP_LACIA	P23496 lactococcus
12	61	13.6	426	1 YG43_ARCFU	O28630 archaeoglob
13	61	13.6	599	1 CENB_HUMAN	P07199 homo sapien
14	60	13.4	240	1 Y162_HELPJ	Q9zmr9 helicobacte
15	60	13.4	240	1 Y162_HELPJ	O4970 helicobacte
16	60	13.4	613	1 CGAA_CLOBI	Q45882 clostridium
17	59.5	13.3	363	1 R23A_HUMAN	P54725 homo sapien
18	59	13.2	251	1 FLGF_SALTY	P16323 salmonella
19	59	13.2	329	1 Y429_CHLTR	O84436 chlamydia t
20	59	13.2	520	1 CLIA_PIG	P10612 sus scrofa
21	59	13.2	2163	1 BRR2_YEAST	P32639 saccharomyc
22	58.5	13.1	270	1 YKAA_CAEEL	P34261 caenorhabdi
23	58	12.9	297	1 YIM4_YEAST	P40471 saccharomyc
24	58	12.9	891	1 DPOL_HELPY	P56105 helicobacte
25	58	12.9	1234	1 PIP3_HUMAN	Q01970 homo sapien
26	57.5	12.8	213	1 ATPO_BOVIN	P13621 bos taurus
27	57.5	12.8	331	1 FLGI_THEMA	Q9xlm5 thermotoga
28	57.5	12.8	363	1 R23A_MOUSE	P54726 mus musculu
29	57.5	12.8	475	1 U2AF_HUMAN	P26368 homo sapien
30	57.5	12.8	536	1 MVIN_CHLMU	Q9pjb9 chlamydia m
31	57.5	12.8	638	1 YQGS_BAGSU	P54496 bacillus su
32	57.5	12.8	875	1 AMD2_XENIA	P12890 xenopus lae
33	57.5	12.8	2292	1 POLG_EMCVB	P17593 encephalomy

RESULT 1

ID	YFL4_YEAST	STANDARD	PRT	481 AA
AC	P43616;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Hypothetical 52.9 kDa protein in SAPI55-YMR31 intergenic region.			
GN	YFR044C.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C / AB972;			
RX	MEDLINE=96287652; PubMed=7670463;			
RA	Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,			
RA	Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,			
RA	Yamazaki M., Tashiro H., Eki T.;			
RT	"Analysis of the nucleotide sequence of chromosome VI from Saccharomyces cerevisiae.";			
RL	Nat. Genet. 10:261-268(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C / AB972;			
RX	MEDLINE=96287652; PubMed=8686379;			
RA	Eki T., Naitou M., Hagiwara H., Shibata T., Hanaoka F., Murakami Y.;			
RA	Sasanuma M., Tsuchiya Y., Shibata T.,			
RT	"Analysis of a 36.2 kb DNA sequence including the right telomere of chromosome VI from Saccharomyces cerevisiae.";			
RL	Yeast 12:149-167(1996).			
CC	- - SIMILARITY: STRONG, TO THE C-TERMINAL HALF OF YEAST YBR281C.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; D50617; BAA09283.1; -			
DR	EMBL; D44597; BAA08010.1; -			
DR	SGD; S0001940; YFR044C.			
DR	InterPro: IPR002933; Peptidase_M20.			
DR	Pfam; PF01546; Peptidase_M20; 1.			
KW	Hypothetical protein.			
SQ	SEQUENCE 481 AA; 52871 MW; 3E53773A945F5EBC CRC64;			

Query Match 16.3%; Score 73; DB 1; Length 481;  
Best Local Similarity 28.6%; Pred. No. 0.99;  
Matches 28; Conservative 11; Mismatches 31; Indels 28; Gaps 4;  
Qy 17 KTLILLDVSTKNPVRTV-----NENFSL-----QLDPSIIHDG--WL-- 52







```
FT METAL 418 418 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 442 442 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 457 457 POTENTIAL.
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 617 617 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 32 32 S -> P.
FT VARIANT 49 49 /FTID=VAR_011710.
FT VARIANT 712 712 S -> G.
FT VARIANT 712 712 /FTID=VAR_011711.
FT VARIANT 712 712 R -> S.
FT SEQUENCE 732 AA; 83330 MW; 80E0D19CFA642313 CRC64;

Query Match 14.0%; Score 62.5; DB 1; Length 732;
Best Local Similarity 32.4%; Pred. No. 26;
Matches 22; Conservative 8; Mismatches 19; Indels 19; Gaps 3;

QY 2 DRRPLPVDRAAGLKERTLLIDVSTKNPRTV-NENFLSLQLDPSIIH----- 48
Db 154 ERAALP---AOELEENKILLDMETTSYVATVCHPGNSCLQLEPDLTNWMTSRKYEDLL 210
QY 49 ---DGLWD 53
Db 211 WAVEGWRD 218

RESULT 8
ID HGBB_HAEBIN STANDARD; PRT; 1067 AA.
AC Q9KIV1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin and hemoglobin-haptoglobin binding protein B precursor
DE (Hemoglobin binding protein B).
GN HGBB.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTHI N182;
RX MEDLINE=20316037; PubMed=10858226;
RA Cope L.D., Hrkai Z., Hansen E.J.;
RT "Detection of phase variation in expression of proteins involved in
hemoglobin and hemoglobin-haptoglobin binding by nontypeable
Haemophilus influenzae.";
RL Infect. Immun. 68:4092-4101(2000).
CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED
FOR HEME UPTAKE.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA
REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAGING.
ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A
WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.
CC -!- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
```

```
CC entties requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AF221059; AAF80177.1; -.
CC InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC.1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Transport; TonB box; Multigene family; Signal;
KW Receptor; Repeat.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1067 HEMOGLOBIN AND HEMOGLOBIN-HAPTOGLOBIN
BINDING PROTEIN B.
FT DOMAIN 26 49 6 X 4 AA TANDEM REPEATS OF Q-P-T-N.
FT REPEAT 26 29 1.
FT REPEAT 30 33 2.
FT REPEAT 34 37 3.
FT REPEAT 38 41 4.
FT REPEAT 42 45 5.
FT REPEAT 46 49 6.
FT SITE 59 66 TONB BOX.
FT SITE 1050 1067 TONB C-TERMINAL BOX.
FT SEQUENCE 1067 AA; 122471 MW; 29D295DC4747632E CRC64;

Query Match 14.0%; Score 62.5; DB 1; Length 1067;
Best Local Similarity 27.8%; Pred. No. 40;
Matches 27; Conservative 18; Mismatches 33; Indels 19; Gaps 5;

QY 1 GDRRLPVDRAAGLKERTLI-----LLDVSTKNPRTVNEN-----FLSLQLDPSI 46
Db 598 GEKKRPDISVAGCANGTLLHSDIGDKTYLIPVTKNNVLYFGDNVOLTSLWGLDLNRY 657
QY 47 IHGWLDFLSSKRLVTLARGL-SPAFLRFGGKRTDFL 82
Db 658 DHVKYLP--SYDKNIPVPKGLITGLFKKF--KSTDIV 690

RESULT 9
ID ACE_HUMAN STANDARD; PRT; 1306 AA.
AC P12821;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Angiotensin-converting enzyme, somatic isoform precursor (EC 3.4.15.1)
DE (ACE) (Dipeptidyl carboxypeptidase I) (Kinase II) (CD143 antigen).
GN DCP1 OR DCP OR ACE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89071703; PubMed=2849100;
RA Soubrier F., Alhenc-Gelas F., Hubert C., Allegrini J., John M.,
Tregear G., Corbol P.;
RT "Two putative active centers in human angiotensin I-converting enzyme
revealed by molecular cloning.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9386-9390(1988).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANTS S-261; W-561 AND S-1286.
RX MEDLINE=99251580; PubMed=10319862;
RA Rieder M.J., Taylor S.L., Clark A.G., Nickerson D.A.;
RT "Sequence variation in the human angiotensin converting enzyme.";
RL Nat. Genet. 22:59-62(1999).
RN [3]
RP PARTIAL SEQUENCE OF 30-46.
RC TISSUE=Lung.
RX MEDLINE=90110025; PubMed=2558109;
RA Takeuchi K., Shimizu T., Ohishi N., Seyama Y., Takaku F.,
Yotsumoto H.;
RT "Purification of human lung angiotensin-converting enzyme by high-
```

performance liquid chromatography: properties and N-terminal amino acid sequence.";

J. Biochem. 106:442-445(1989).

[4]

ZINC-BINDING.

Medline=91308093; PubMed=1649623;

Enlers M.R., Riordan J.F.;

"Angiotensin-converting enzyme: zinc- and inhibitor-binding stoichiometries of the somatic and testis isozymes.";

Biochemistry 30:7118-7126(1991).

CC -1- FUNCTION: CONVERTS ANGIOTENSIN I TO ANGIOTENSIN II BY RELEASE OF THE TERMINAL HIS-LEU, THIS RESULTS IN AN INCREASE OF THE VASOCONSTRICTOR ACTIVITY OF ANGIOTENSIN.

CC -1- CATALYTIC ACTIVITY: Release of a C-terminal dipeptide, oligopeptide-l-xaa-xbb, when xaa is not pro, and xbb is neither asp nor glu. Converts angiotensin I to angiotensin II.

CC -1- COFACTOR: BINDS 2 ZINC IONS.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- ALTERNATIVE PRODUCTS: THE TESTICULAR ANGIOTENSIN-CONVERTING ENZYME IS TRANSCRIBED FROM THE SAME GENE AS THE SOMATIC ISOFORM, PROBABLY FROM AN ALTERNATIVE START SITE.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M2.

CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD143 entry; WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd143.htm".

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; J04144; AAA51684.1; .

DR EMBL; AF118569; AAD28560.1; .

DR PIR; A31759; A31759.

DR PIR; PQ0004; PQ0004.

DR MIM; 106180; .

DR InterPro; IPR0001548; Peptidase\_M2.

DR InterPro; IPR000130; Zn\_Mrpeptidase.

DR Pfam; PF01401; Peptidase\_M2; 2.

DR PRINTS; PR00791; PEPTIDPASEA.

DR ProDom; PD004184; Peptidase\_M2; 2.

DR PROSITE; PS00142; ZINC\_PROTEASE; 2.

DR HydroLase; Metalloprotease; Zinc; Dipeptidase;

KW Glycoprotein; Transmembrane; Repeat; Signal; Alternative splicing;

KW Polymorphism.

FT SIGNAL 1 29

FT CHAIN 30 1306

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394 394

FT METAL 988 988

FT ACT\_SITE 989 989

FT METAL 992 992

FT CARBOHYD 38 38

FT CARBOHYD 54 54

FT CARBOHYD 74 74

FT CARBOHYD 79 79

FT CARBOHYD 111 111

FT CARBOHYD 146 146

FT CARBOHYD 160 160

FT CARBOHYD 272 272

FT CARBOHYD 318 318

FT CARBOHYD 445 445

FT CARBOHYD 509 509

FT CARBOHYD 677 677

FT DOMAIN 30 1259

FT TRANSMEM 1260 1276

FT DOMAIN 1277 1306

FT REPEAT 227 583

FT REPEAT 825 1181

FT METAL 390 390

FT ACT\_SITE 391 391

FT METAL 394



```
QY 79 TDFLQFQNL 87
   || | ||
Db 236 VDFDFPNL 244

RESULT 11
LAXP_LACLA STANDARD; PRT; 299 AA.
AC P23496;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Lax protein, plasmid.
GN LAX.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OG Plasmid pmg820.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91093107; PubMed=2125052;
RA de Vos W.M., Boerrigter I., van Rooyen R.J., Reiche B.,
RA Hengstenberg W.;
RT "Characterization of the lactose-specific enzymes of the
RT phosphotransferase system in Lactococcus lactis.";
RL J. Biol. Chem. 265:22554-22560(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60447; AAA25184.1; -.
DR PIR; D23696; D23696.
DR InterPro; IPR003843; UPF0010_DUF.
DR Pfam; PF02693; UPF0010; 1.
KW Plasmid.
SQ SEQUENCE 299 AA; 34519 MW; 715599C040DCD608 CRC64;

Query Match 13.7%; Score 61.5; DB 1; Length 299;
Best Local Similarity 42.0%; Pred. No. 11;
Matches 29; Conservative 3; Mismatches 26; Indels 11; Gaps 5;

QY 21 LLDVSTKNPVRTNENFLSLQDPSII-HDG-WLDFLSSKRLVTLARGLSPAFLRFGGKR 78
   ||| : | || || | : || || || | || |
Db 185 LLDLQDRTPL--ENOKSLDLDYSLFSDAITLDRKS-RSVTLRSRKGK-----GLR 235

QY 79 TDFLQFQNL 87
   || | ||
Db 236 VDFDFPNL 244

RESULT 12
YG43_ARCFU STANDARD; PRT; 426 AA.
AC O28630;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF1643.
GN AF1643.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyripides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., Mckenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -i- SIMILARITY: BELONGS TO THE UBID FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000989; AAB89599.1; -.
DR TIGR; AF1643; -.
DR InterPro; IPR002830; UPF0096.
DR Pfam; PF01977; UPF0096; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 426 AA; 47715 MW; 7EBCE276F501E2D8 CRC64;

Query Match 13.6%; Score 61; DB 1; Length 426;
Best Local Similarity 29.9%; Pred. No. 20;
Matches 23; Conservative 7; Mismatches 21; Indels 26; Gaps 4;

QY 13 GLK-EKTLILID-----VSTKNPVRTNENFLSLQDPSIIHDGWLDFSSKRLV 61
   ||| : | || || | : || || || | || |
Db 2 GIKTKNFICLKRTFSMNLRAISVANPV-----QLEEEIKHDEVVSLKSNLL 51

QY 62 TLARGLSPAFLRFGGKR 78
   | | ||
Db 52 D-----KPVILNVEGKK 63

RESULT 13
CENB_HUMAN STANDARD; PRT; 599 AA.
ID CENB_HUMAN
AC P07199;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Major centromere autoantigen B (Centromere protein B) (CENP-B).
GN CENPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91372020; PubMed=1893793;
RA Sullivan K.F., Glass C.A.;
RT "CENP-B is a highly conserved mammalian centromere protein with
RT homology to the helix-loop-helix family of proteins.";
RL Chromosoma 100:360-370(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
```



```
RESULT 15
Y162_HELPY
ID Y162_HELPY STANDARD; PRT; 240 AA.
AC O24970;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein HP0162.
GN HP0162.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G., B.A.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty S.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -!- SIMILARITY: BELONGS TO THE UPF0082 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000537; AAD07231.1; -.
CC TIGR; HP0162; -.
DR InterPro; IPR002876; DUF28.
DR Pfam; PF01709; DUF28; 1.
DR Prodom; PD004323; DUF28; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 240 AA; 27127 MW; 6BB7E0DC87B35F2E CRC64;

Query Match 13.4%; Score 60; DB 1; Length 240;
Best Local Similarity 28.9%; Pred. No. 13;
Matches 13; Conservative 15; Mismatches 15; Indels 2; Gaps 1;

QY 19 LILLDVSTKNPVTN--ENFLSLQLDPSIIHDGWLDFLSKRLV 61
|:::| | | |:::| | | |:::|
DB 92 LIIMECHTDNPTRIANLKSIFYNKYTGASIVPNGSLEFMNRKSV 136
```

Search completed: July. 30, 2002, 08:31:27  
Job time: 974 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:30:41 ; Search time 123.94 Seconds  
(without alignments)  
122.830 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_42\_129  
Perfect score: 448  
Sequence: 1 GDRPLPVDRAAGLKEKTLI.....PAFLRFGGKRTDFLOFQNL 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	448	100.0	480	4 Q9HB39	Q9hb39 homo sapien
2	448	100.0	534	4 Q9HB38	Q9hb38 homo sapien
3	448	100.0	592	4 Q9HB37	Q9hb37 homo sapien
4	155	34.6	536	11 Q9QZF8	Q9qzf8 rattus norv
5	151.5	33.8	545	4 Q9UL39	Q9ul39 homo sapien
6	150.5	33.6	543	4 Q9Y251	Q9y251 homo sapien
7	140.5	31.4	545	6 Q9MY0	Q9my0 bos taurus
8	132.5	29.6	523	13 Q9YK5	Q9yk5 gallus gall
9	72.5	16.2	304	10 Q9C5M5	Q9c5m5 arabidopsis
10	72.5	16.2	334	10 Q9LYF5	Q9lyf5 arabidopsis
11	70.5	15.7	527	10 Q9LRC8	Q9lrc8 scutellaria
12	68.5	15.3	329	10 Q9FG13	Q9fg13 arabidopsis
13	68.5	15.3	694	16 Q98PE1	Q98pe1 mycoplasma
14	66	14.7	305	12 Q9LMM5	Q9lmm5 lumby skin
15	65.5	14.6	1391	5 Q1772	Q1772 caenorhabdi
16	65	14.5	921	2 Q9EYGS	Q9eygs actinobacil

17	65	14.5	2288	3	060033	060033 emericella
18	64.5	14.4	199	5	Q9BI00	Q9bi00 globodera p
19	64.5	14.4	702	10	Q9FLL6	Q9fil6 arabidopsis
20	64.5	14.4	881	5	Q9GR11	Q9gr11 paramecium
21	64.5	14.4	2126	2	Q93LR2	Q93lr2 microcystis
22	64.5	14.4	2126	2	Q93LR1	Q93lr1 microcystis
23	64	14.3	369	10	Q42606	Q42606 arabidopsis
24	64	14.3	536	10	Q9C825	Q9c825 arabidopsis
25	64	14.3	1477	3	Q74218	Q74218 candida alb
26	63.5	14.2	497	5	Q18652	Q18652 caenorhabdi
27	63.5	14.2	756	16	Q988J0	Q988j0 rhizobium 1
28	63.5	14.2	935	5	Q9VE79	Q9ve79 drosophila
29	63.5	14.2	2404	5	Q9VE34	Q9ve34 drosophila
30	62.5	14.0	326	16	Q97DT6	Q97dt6 clostridium
31	62.5	14.0	369	10	Q96532	Q96532 arabidopsis
32	62.5	14.0	694	4	Q15540	Q15540 homo sapien
33	62.5	14.0	732	6	Q9GLN6	Q9gln6 pan troglod
34	62.5	14.0	1304	6	Q9GLN7	Q9gln7 pan troglod
35	62.5	14.0	2126	2	Q9SIA8	Q9sla8 microcystis
36	62	13.8	144	6	Q9TU38	Q9t38 oryctolagus
37	62	13.8	292	10	Q9LG45	Q9lg45 oryza sativ
38	62	13.8	295	16	Q98H10	Q98h10 rhizobium 1
39	62	13.8	749	4	Q60741	Q60741 homo sapien
40	62	13.8	822	6	Q9MZ51	Q9mz51 oryctolagus
41	62	13.8	830	16	Q9KGI3	Q9kgi3 bacillus ha
42	62	13.8	910	11	Q54899	Q54899 mus musculu
43	62	13.8	910	11	Q9JKB0	Q9jkb0 rattus norv
44	62	13.8	910	11	Q88704	Q88704 mus musculu
45	61.5	13.7	126	17	Q57867	Q57867 pyrococcus

ALIGNMENTS

RESULT 1

Q9HB39 PRELIMINARY; PRT; 480 AA.  
ID Q9HB39  
AC Q9HB39;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE HEPARANASE-LIKE PROTEIN HPA2A.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20483645; PubMed=11027606;  
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,  
RA Hircok M., Patel S., Barry E., Stubbfield C., Terrett J., Page M.;  
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian  
RT Heparanase Family Member."  
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).  
DR EMBL; AF282885; AAC23421.1; -;  
SQ SEQUENCE 480 AA; 53900 MW; F75F89F67AC1FF83 CRC64;

Query Match 100.0%; Score 448; DB 4; Length 480;  
Best Local Similarity 100.0%; Pred. No. 3.5e-43;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIIHDGWLDFLSKRL 60  
|||||  
Db 42 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQLDPSIIHDGWLDFLSKRL 101  
|||||

Qy 61 VTLAGLSPAFLRFGGKRTDFLOFQNL 88  
|||||

Db 102 VTLAGLSPAFLRFGGKRTDFLOFQNL 129  
|||||

RESULT 2

Q9HB38

ID Q9HB38 PRELIMINARY; PRT; 534 AA.  
AC Q9HB38;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE HEPARANASE-LIKE PROTEIN HPA2B.  
DE HEPARANASE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20483645; PubMed=11027606;  
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,  
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;  
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian  
RT Heparanase Family Member.";  
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).  
DR EMBL; AF282886; AAG23422.1; -.  
SQ SEQUENCE 534 AA; 60063 MW; C3DE5E900CB338C4 CRC64; 0

Query Match 100.0%; Score 448; DB 4; Length 534;  
Best Local Similarity 100.0%; Pred. No. 3.9e-43;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIHDGWLDFLSSKRL 60  
DB 42 GDRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIHDGWLDFLSSKRL 101  
QY 61 VTLARGLSPAFLRGGRKRTDFLQFNLR 88  
DB 102 VTLARGLSPAFLRGGRKRTDFLQFNLR 129

RESULT 3  
Q9HB37  
ID Q9HB37 PRELIMINARY; PRT; 592 AA.  
AC Q9HB37;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE HEPARANASE-LIKE PROTEIN HPA2C.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20483645; PubMed=11027606;  
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,  
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;  
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian  
RT Heparanase Family Member.";  
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).  
DR EMBL; AF282887; AAG23423.1; -.  
SQ SEQUENCE 592 AA; 66580 MW; 95C384AD9A74258E CRC64;

Query Match 100.0%; Score 448; DB 4; Length 592;  
Best Local Similarity 100.0%; Pred. No. 4.5e-43;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIHDGWLDFLSSKRL 60  
DB 42 GDRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIHDGWLDFLSSKRL 101  
QY 61 VTLARGLSPAFLRGGRKRTDFLQFNLR 88  
DB 102 VTLARGLSPAFLRGGRKRTDFLQFNLR 129

RESULT 4

Q9QZF8  
ID Q9QZF8 PRELIMINARY; PRT; 536 AA.  
AC Q9QZF8;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HEPARANASE.  
GN HEP.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Podyna K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;  
RT "Heparanase from parathyroid cell line.";  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF184967; AAF04563.1; -.  
SQ SEQUENCE 536 AA; 60569 MW; 6208B1FD9EE28421 CRC64;

Query Match 34.6%; Score 155; DB 11; Length 536;  
Best Local Similarity 46.3%; Pred. No. 2e-09;  
Matches 40; Conservative 13; Mismatches 31; Indels 2; Gaps 2;  
QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIHDGWLDFLSSKRL 58  
DB 12 GRURALTQGTAGTAPTKDVVDLEFYTKRLQFQSVSPFLSTIDASLATDPRFLTLGSP 71  
QY 59 RLVTARGLSPAFLRGGRKRTDFLQF 84  
DB 72 RLVTARGLSPAFLRGGRKRTDFLQF 97

RESULT 5  
Q9UL39  
ID Q9UL39 PRELIMINARY; PRT; 545 AA.  
AC Q9UL39;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HEPARANASE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=PLACENTA;  
RX MEDLINE=20229546; PubMed=10764835;  
RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;  
RT "Heparanase expression in invasive trophoblasts and acute vascular  
RT damage.";  
RL Glycobiology 10:467-475(2000).  
DR EMBL; AF084467; AAD54516.1; -.  
SQ SEQUENCE 545 AA; 61418 MW; 67B80ACD73C5A9A1 CRC64;

Query Match 33.8%; Score 151.5; DB 4; Length 545;  
Best Local Similarity 47.1%; Pred. No. 5.2e-09;  
Matches 33; Conservative 16; Mismatches 20; Indels 1; Gaps 1;  
QY 16 EKTLLLDVSTKNPVRTVNENFLSLQLDPSIHDGWLDFLSSKRLVTARGLSPAFLR 74  
DB 37 QQDVVDLDFTQEPHLVSPFLSVTDANLATDPRFLILGSPKRLTARGLSPAYLRF 96  
QY 75 GGRKRTDFLQF 84  
DB 97 GGRKRTDFLQF 106

RESULT 6  
Q9Y251

ID Q9Y251 PRELIMINARY; PRT; 543 AA.  
AC Q9Y251;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE HEPARANASE.  
GN HPA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RX MEDLINE=993321249; PubMed=10395326;  
RA Hulett M.J., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,  
RA Parish C.R.;  
RT "Cloning of mammalian heparanase, an important enzyme in tumor  
RT invasion and metastasis";  
RL Nat. Med. 5:803-809(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Vlodavsky I., Friedman Y., Elkin M., Aingorn H., Atzmon R.,  
RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,  
RA Spector L., Pecker I.;  
RT "Mammalian heparanase: a novel gene involved in tumor progression and  
RT metastasis";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99377052; PubMed=10446189;  
RA Toyoshima M., Nakajima M.;  
RT "Human heparanase. Purification, characterization, cloning, and  
RT expression";  
RL J. Biol. Chem. 274:24153-24160(1999).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RX MEDLINE=99335379; PubMed=10405343;  
RA Kussie P.H., Hulmes J.D., Ludwig D., Patel S., Navarro E.C.,  
RA Seddon A.P., Giorgio N.A., Bohlen P.;  
RT "Cloning and Functional Expression of a Human Heparanase Gene";  
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).  
DR EMBL; AF165154; AAD45379.1; -  
DR EMBL; AF144325; AAD41342.1; -  
DR EMBL; AF155510; AAD54941.1; -  
DR EMBL; AF152376; AAD45669.1; -  
SQ SEQUENCE 543 AA; 61176 MW; AD262EC267334AB2 CRC64;

Query Match 33.6%; Score 150.5; DB 4; Length 543;  
Best Local Similarity 43.5%; Pred. No. 6.7e-09;  
Matches 37; Conservative 17; Mismatches 28; Indels 3; Gaps 2;

QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQDPSIHD-GWLDPLSKR 59  
Db 22 GPLSPGALPRA--QAQDVNDLDFEFTQPLHVSFSLVTDANLATDPRFLILGSPK 79  
QY 60 LVTLARGLSPAFRLFGGKRTDFLOF 84  
Db 80 LRTLARGLSPAYLRFGGTKDTDFLIF 104

RESULT 7  
Q9MYO0 PRELIMINARY; PRT; 545 AA.  
AC Q9MYO0;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE HEPARANASE.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

ID Q9Y251 PRELIMINARY; PRT; 543 AA.  
AC Q9Y251;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE HEPARANASE.  
GN HPA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RX MEDLINE=993321249; PubMed=10395326;  
RA Hulett M.J., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,  
RA Parish C.R.;  
RT "Cloning of mammalian heparanase, an important enzyme in tumor  
RT invasion and metastasis";  
RL Nat. Med. 5:803-809(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Vlodavsky I., Friedman Y., Elkin M., Aingorn H., Atzmon R.,  
RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,  
RA Spector L., Pecker I.;  
RT "Mammalian heparanase: a novel gene involved in tumor progression and  
RT metastasis";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99377052; PubMed=10446189;  
RA Toyoshima M., Nakajima M.;  
RT "Human heparanase. Purification, characterization, cloning, and  
RT expression";  
RL J. Biol. Chem. 274:24153-24160(1999).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RX MEDLINE=99335379; PubMed=10405343;  
RA Kussie P.H., Hulmes J.D., Ludwig D., Patel S., Navarro E.C.,  
RA Seddon A.P., Giorgio N.A., Bohlen P.;  
RT "Cloning and Functional Expression of a Human Heparanase Gene";  
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).  
DR EMBL; AF165154; AAD45379.1; -  
DR EMBL; AF144325; AAD41342.1; -  
DR EMBL; AF155510; AAD54941.1; -  
DR EMBL; AF152376; AAD45669.1; -  
SQ SEQUENCE 543 AA; 61176 MW; AD262EC267334AB2 CRC64;

Query Match 33.6%; Score 150.5; DB 4; Length 543;  
Best Local Similarity 43.5%; Pred. No. 6.7e-09;  
Matches 37; Conservative 17; Mismatches 28; Indels 3; Gaps 2;

QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFLSLQDPSIHD-GWLDPLSKR 59  
Db 22 GPLSPGALPRA--QAQDVNDLDFEFTQPLHVSFSLVTDANLATDPRFLILGSPK 79  
QY 60 LVTLARGLSPAFRLFGGKRTDFLOF 84  
Db 80 LRTLARGLSPAYLRFGGTKDTDFLIF 104

RESULT 7  
Q9MYO0 PRELIMINARY; PRT; 545 AA.  
AC Q9MYO0;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE HEPARANASE.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RA Kizaki K., Nakano H., Takahashi T., Imai K., Hashizume K.;  
RT "Expression of Heparanase mRNA in Bovine Placenta During Gestation.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF281160; AAF87301.2; -  
SQ SEQUENCE 545 AA; 61076 MW; FAC4BDFD855B933 CRC64;

Query Match 31.4%; Score 140.5; DB 6; Length 545;  
Best Local Similarity 43.0%; Pred. No. 9.5e-08;  
Matches 34; Conservative 12; Mismatches 24; Indels 9; Gaps 2;

QY 7 PVDRAGLKEKTLILLDVSTKNPVTNENFLSLQDPSIHD-GWLDPLSKRLVTLAR 65  
Db 36 PADDAAE-----LEFFTERPLHVSFAFLSFTIDANLATDPRFFTLGSSKLRTLAR 87  
QY 66 GLSPAFRLFGGKRTDFLOF 84  
Db 88 GLAPAYLRFGGKNGDFLIF 106

RESULT 8  
Q90YK5 PRELIMINARY; PRT; 523 AA.  
ID Q90YK5;  
AC Q90YK5;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HEPARANASE.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=11387326;  
RA Goldsmith O., Zcharia E., Aingorn H., Guatta-Rangini Z., Atzmon R.,  
RA Michal I., Pecker I., Mitrani E., Vlodavsky I.;  
RT "Expression Pattern and Secretion of Human and Chicken Heparanase Are  
RT Determined by Their Signal Peptide Sequence.";  
RL J. Biol. Chem. 276:29178-29187(2001).  
DR EMBL; AY037007; AAK82648.1; -  
SQ SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF881 CRC64;

Query Match 29.6%; Score 132.5; DB 13; Length 523;  
Best Local Similarity 48.4%; Pred. No. 7.6e-07;  
Matches 31; Conservative 8; Mismatches 24; Indels 1; Gaps 1;

QY 22 LDVSTKNPVTNENFLSLQDPSIHD-GWLDPLSKRLVTLARGLSPAFRLFGGKRTD 80  
Db 21 LQLGLREFIGAVSPAFSLSLTLDASLARDPRFVALLRHKPLHTLASGLSPGLRFGGTSTD 80  
QY 81 FLQF 84  
Db 81 FLIF 84

RESULT 9  
Q9C5M5 PRELIMINARY; PRT; 304 AA.  
ID Q9C5M5;  
AC Q9C5M5;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 33.0 KDA PROTEIN (AT1G27000/T7N9\_6).  
GN T7N9.6.

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 ON NCBI\_taxid=3702;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,  
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RT "Full length cDNA of gene T7N9.6 (GI:8778852).";  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.;  
 RT "Arabidopsis cDNA clones";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banh J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,  
 RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,  
 RA Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,  
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,  
 RA Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,  
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,  
 RA Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RT "Full length cDNA of gene T7N9.6 (GI:8778852).";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF360143; AA25853.1;  
 DR EMBL; AY057525; AAL09765.1;  
 DR EMBL; AY056367; AAL07253.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 304 AA; 33013 MW; 463DB4579AF52195 CRC64;

Query Match 16.28; Score 72.5; DB 10; Length 304;  
 Best Local Similarity 31.08; Pred. No. 3.2; Indels 9; Gaps 3;  
 Matches 27; Conservative 15; Mismatches 36;  
 QY 10 RAAGLKEKTLILLDVSTK--NPVRTVNENFLSLQDPSIIH-----DGWLDFLSSKRLV 61  
 Db 157 RIONLDDKVEKIDLSKINSQVISARENISSEMDLESNLHNLITGLDGKLDLTLEYKQDV 216  
 QY 62 TIARGLSPAFLRFGGKRTDFLQFNLR 88  
 Db 217 TNVFMNL-LYNVFGGKSTKLPEMEQLQ 242  
 RESULT 10  
 ID Q9LFY5 PRELIMINARY; PRT; 334 AA.  
 AC Q9LFY5;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE T7N9.6.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 ON NCBI\_taxid=3702;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,  
 RA Shinn P., Altafi H., Bei B., Chin C., Chlou J., Choi E., Conn L.,  
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,  
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
 RA Ecker J.R.;  
 RT "Genomic sequence for Arabidopsis thaliana BAC T7N9 from chromosome I";  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chlou J., Choi E.,  
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,  
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
 RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
 RA Theologis A., Ecker J.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chlou J., Choi E.,  
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,  
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
 RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
 RA Theologis A., Ecker J.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC000348; AAF79851.1;  
 SQ SEQUENCE 334 AA; 36323 MW; 82391C5498EF3760 CRC64;

Query Match 16.28; Score 72.5; DB 10; Length 334;  
 Best Local Similarity 31.08; Pred. No. 3.5; Indels 9; Gaps 3;  
 Matches 27; Conservative 15; Mismatches 36;  
 QY 10 RAAGLKEKTLILLDVSTK--NPVRTVNENFLSLQDPSIIH-----DGWLDFLSSKRLV 61  
 Db 163 RIONLDDKVEKIDLSKINSQVISARENISSEMDLESNLHNLITGLDGKLDLTLEYKQDV 222  
 QY 62 TIARGLSPAFLRFGGKRTDFLQFNLR 88  
 Db 223 TNVFMNL-LYNVFGGKSTKLPEMEQLQ 248  
 RESULT 11  
 ID Q9LRC8 PRELIMINARY; PRT; 527 AA.  
 AC Q9LRC8;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE BETA-GLUCURONIDASE.  
 GN SGUS.  
 OS Scutellaria baicalensis.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;



```
OC Asteridae: euasterids I; Lamiales; Lamiaceae; Scutellaria.
OX NCBI_TaxID=65409;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=204118130; PubMed=10858442;
RA Sasaki K., Taura F., Shoyama Y., Morimoto S.;
RT "Molecular Characterization of a Novel beta-Glucuronidase from
RT Scutellaria baicalensis Georgi.";
RL J. Biol. Chem. 275:27466-27472(2000).
DR EMBL: AB040072; BAA97804.1; -
DR InterPro: IPR001179; FKBP_PPIase.
DR PROSITE: PS00453; FKBP_PPIASE.1; UNKNOWN.1.
SQ SEQUENCE 527 AA; 58772 MW; ASDE7C423FAIE2B CRC64;

Query Match 15.3%; Score 70.5; DB 10; Length 527;
Best Local Similarity 28.7%; Pred. No. 10;
Matches 25; Conservative 17; Mismatches 28; Indels 17; Gaps 5;

QY 15 KETLLILDVSTKNPVTNENFLSLQD---PSIIHDG---W-----LDFSSKRLV 61
Db 26 EETIVKIE---ENPVAQTIDENYVCATLDLWPTKCYGNCPCWCKSSFLNLD-LNNNIIR 81
QY 62 TLARGLSPAFLRFGGKRTDFLOFNLR 88
Db 82 NAVKEFAPLKLRFGGTLDRLVYQTSR 108

RESULT 12
Q9FG13 ID Q9FG13 PRELIMINARY; PRT; 329 AA.
AC Q9FG13;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GB|A04946.2;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Katoh T., Asanizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP002543; BAB11406.1; -
DR InterPro: IPR000379; Est_lip_thioest_actsite.
SQ SEQUENCE 329 AA; 36597 MW; 000140D8BB91126A CRC64;

Query Match 15.3%; Score 68.5; DB 10; Length 329;
Best Local Similarity 25.4%; Pred. No. 10;
Matches 15; Conservative 14; Mismatches 25; Indels 5; Gaps 1;

QY 28 NPVTNENFLSLQDPSIIHDGWLDFSSKRLVTLARGLSPAFLRFGGKRTDFLOFN 86
Db 246 NPFGPTPTLESISLEPMLVVGSGELLRDR-----AKEAYKLKMGKRVDFEEN 299

RESULT 13
Q98PE1 ID Q98PE1 PRELIMINARY; PRT; 694 AA.
AC Q98PE1;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE EXODEOXYRIBONUCLEASE V ALPHA CHAIN (EC 3.1.11.5).
GN MYPU.7820.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
```

```
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=UAB CTIP;
RX MEDLINE=21267165; PubMed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
RT Mycoplasma pulmonis.";
RL Nucleic Acids Res. 29:2145-2153(2001).
DR EMBL: AL445365; CAC13955.1; -
DR MypuList; MYPU.7820; -
DR InterPro: IPR003975; Shal_channel.
DR PRINTS: PR01497; SHALCHANNEL.
KW Hydrolase; Complete proteome.
KW SEQUENCE 694 AA; 80774 MW; 61D38974D4465132 CRC64;

Query Match 15.3%; Score 68.5; DB 16; Length 694;
Best Local Similarity 35.4%; Pred. No. 24;
Matches 17; Conservative 8; Mismatches 16; Indels 7; Gaps 1;

QY 1 GRRRLPVDRAAGLKEKTLILLDVST-----KNPVTNENFLSLQ 41
Db 428 GDRQLPAIGAGNLLDLLELQNISVNLRSQIFRNKTKDISENFLVQ 475

RESULT 14
Q91MM5 ID Q91MM5 PRELIMINARY; PRT; 305 AA.
AC Q91MM5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LSDV139 PUTATIVE SER/THR PROTEIN KINASE.
GN LSDV139.
OS lumpy skin disease virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Capripoxvirus.
OX NCBI_TaxID=59509;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEETHLING 2490;
RX MEDLINE=21329495; PubMed=11435593;
RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RT "Genome of lumpy skin disease virus.";
RL J. Virol. 75:7122-7130(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NEETHLING 2490;
RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF325528; AAK85100.1; -
KW Kinase.
SQ SEQUENCE 305 AA; 35635 MW; F77A0E94273AB816 CRC64;

Query Match 14.7%; Score 66; DB 12; Length 305;
Best Local Similarity 23.0%; Pred. No. 18;
Matches 14; Conservative 17; Mismatches 16; Indels 14; Gaps 2;

QY 20 ILLDVSTKNPVR-----TVNENFLSLQDPSIIHDGWLDFSSKRLVTLARGLSP 69
Db 163 IATIDNNKNIYLLDYGLSYRPMINGNHVEYKRDPKKMHNGTIETYS-----IDMHKGVS 218
QY 70 A 70
Db 219 S 219

RESULT 15
```

017772  
ID O17772 PRELIMINARY; PRT; 1391 AA.  
AC O17772;  
DT 01-JAN-1998 (Tremblrel. 05, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE F09C3.1 PROTEIN.  
GN F09C3.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_taxID=6239;  
RN [1]  
RN RP SEQUENCE FROM N.A.  
RA Lennard N.;  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RN RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; Pubmed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
RT investigating biology."  
RL Science 282:2012-2018(1998).  
DR EMBL: Z92781; CAB07179.2; .  
DR InterPro; IPR001715; Calponin\_hom.  
DR InterPro; IPR003247; CH\_type.  
DR InterPro; IPR000048; IQ.  
DR InterPro; IPR001936; RasGAP.  
DR InterPro; IPR000593; RasGAP\_C.  
DR Pfam: PF00612; IQ; 2.  
DR ProDom; PD001527; CH\_type; 1.  
DR ProDom; PD008735; RasGAP\_C; 1.  
DR SMART; SM00033; CH; 1.  
DR SMART; SM00015; IQ; 1.  
DR PROSITE; PS50021; CH; 1.  
DR PROSITE; PS50018; RAS\_GTPASE\_ACTIV\_2; 1.  
SQ SEQUENCE 1391 AA; 159210 MW; 794DF9EBFA2E65AC CRC64;

Query Match 14.6%; Score 65.5; DB 5; Length 1391;  
Best Local Similarity 23.5%; Pred. NO. 1.2e+02;  
Matches 27; Conservative 12; Mismatches 41; Indels 35; Gaps 3;  
  
QY 8 VDRAGLKEKTLILLDVSTKNP-VRTVNENFLSLQID----- 43  
Db 485 VEEAFKLAQVVRINKAKGSEPKVKELEKLSFKLDGCVRELLHWYAKRISSELDASKH 544  
  
QY 44 PSIIHDGWLDF-----LSSKRLVTIARGLSPAPFLREGGKRTDFLQFQNL 87  
Db 545 PESSQDFWLDHKFPLGTIYVNTSTSKHSTTSQEPFLHPNPLRWNPQLQDFVAYENY 599

Search completed: July 30, 2002, 08:30:43  
Job time: 995 sec



```

;
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein
; STREET: 2940 Birchtree lane
; CITY: Silver Spring
; STATE: Maryland
; COUNTRY: United States of America
; ZIP: 20906
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; SOFTWARE: Word for Windows version 2.0 converted to
; SOFTWARE: an ASCII file
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/922,170B
; FILING DATE: 2 SEP 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-922-170B-10
;
;
; Query Match 33.6%; Score 150.5; DB 2; Length 543;
; Best Local Similarity 43.5%; Pred. No. 1.3e-11;
; Matches 37; Conservative 17; Mismatches 28; Indels 3; Gaps 2;
;
; QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHD-GWLDFLSSKR 59
; | | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
; Db 22 GPLSPGALPRP--QAQDVVDLDFFTQEPHLVSPSLVTDANLATDPRFLILGSPK 79
; | | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
; QY 60 LVTLAGLSPAFLRFGGKRTDFLOF 84
; | | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
; Db 80 LRTLARGLSPAYLRFGGTKTDFLIF 104
; | | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
;
; RESULT 3
; US-09-071-739B-2
; Sequence 2, Application US/09071739B
; Patent No. 6177545
; GENERAL INFORMATION:
; APPLICANT: Iris Pecker et al.
; TITLE OF INVENTION: HEPARANASE SPECIFIC MOLECULAR PROBES
; TITLE OF INVENTION: AND THEIR USE IN RESEARCH AND MEDICAL
; TITLE OF INVENTION: APPLICATIONS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 20001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,

```

```

;
; OPERATING SYSTEM: Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,739B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/922,180
; FILING DATE: September 2, 1997
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-09-071-739B-2
;
;
; Query Match 33.6%; Score 150.5; DB 4; Length 543;
; Best Local Similarity 43.5%; Pred. No. 1.3e-11;
; Matches 37; Conservative 17; Mismatches 28; Indels 3; Gaps 2;
;
; QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHD-GWLDFLSSKR 59
; | | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
; Db 22 GPLSPGALPRP--QAQDVVDLDFFTQEPHLVSPSLVTDANLATDPRFLILGSPK 79
; | | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
; QY 60 LVTLAGLSPAFLRFGGKRTDFLOF 84
; | | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
; Db 80 LRTLARGLSPAYLRFGGTKTDFLIF 104
; | | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
;
; RESULT 4
; US-09-181-336-13
; Sequence 13, Application US/09181336A
; Patent No. 6242238
; GENERAL INFORMATION:
; APPLICANT: FREEMAN, Craig Gsoffrey
; APPLICANT: PARISH, Richard
; APPLICANT: HANDORF, Brenton James
; APPLICANT: HULETT, Mark Darren
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
; TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR
; FILE REFERENCE: 032505-004
; CURRENT APPLICATION NUMBER: US/09/181,336A
; CURRENT FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: AU PP0062
; EARLIER FILING DATE: 1997-10-28
; EARLIER APPLICATION NUMBER: AU PP0812
; EARLIER FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
;
; US-09-181-336-13
;
; Query Match 33.6%; Score 150.5; DB 4; Length 543;
; Best Local Similarity 43.5%; Pred. No. 1.3e-11;
; Matches 37; Conservative 17; Mismatches 28; Indels 3; Gaps 2;
;
; QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHD-GWLDFLSSKR 59

```

Db 22 GPLSPGALPRPA--QAQDVVLDLFFFTQEPHLVSPSFLSVTIDANLATDPRFLLILGSPK 79  
QY 60 LVTIARGLSPAFLRFGKRTDFLOF 84  
Db 80 LRTIARGLSPAFLRFGGKTDFLIF 104

RESULT 5  
US-09-260-038B-2  
; Sequence 2, Application US/09260038B  
; Patent No. 6348344  
; GENERAL INFORMATION:  
; APPLICANT: Maty Ayal-Hershkovitz et al.  
; TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR  
; EXPRESSING RECOMBINANT HEPARANASE  
; AND METHODS OF PURIFYING SAME  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina  
; STREET: 2001 Jefferson Davis Highway, Suite 207  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: United States of America  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk  
; COMPUTER: Twinhead\* Slimnote-890TX  
; OPERATING SYSTEM: MS DOS version 6.2,  
; SOFTWARE: Word for Windows version 2.0 converted to  
; an ASCII file  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/260,038B  
; FILING DATE: 02-Mar-1999  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/071,618  
; FILING DATE: May 1, 1998  
; APPLICATION NUMBER: 09/071,739  
; FILING DATE: May 1, 1998  
; APPLICATION NUMBER: 08/922,180  
; FILING DATE: September 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Friedman, Mark M.  
; REGISTRATION NUMBER: 33,883  
; REFERENCE/DOCKET NUMBER: 910/16  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 972-3-562553  
; TELEFAX: 972-3-562554  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 543  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-260-038B-2

Query Match 33.6%; Score 150.5; DB 4; Length 543;  
Best Local Similarity 43.5%; Pred. No. 1.3e-11;  
Matches 37; Conservative 17; Mismatches 28; Indels 3; Gaps 2;  
QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQDPSIIHD-GWLDFLSKR 59  
Db 22 GPLSPGALPRPA--QAQDVVLDLFFFTQEPHLVSPSFLSVTIDANLATDPRFLLILGSPK 79  
QY 60 LVTIARGLSPAFLRFGKRTDFLOF 84  
Db 80 LRTIARGLSPAFLRFGGKTDFLIF 104

RESULT 6  
US-08-481-626-2  
; Sequence 2, Application US/08481626  
; Patent No. 5801040  
; GENERAL INFORMATION:  
; APPLICANT: Soubrier, Florent  
; APPLICANT: Albenc-Gelas, Francois  
; APPLICANT: Hubert, Christine  
; APPLICANT: Corvol, Pierre  
; TITLE OF INVENTION: Nucleic Acid Coding for the Human  
; Testicular Angiotensin Converting Enzyme (ACE) and its  
; Uses, Especially for the In Vitro Screening for this  
; Enzyme in the Organism  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,626  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/656,183  
; FILING DATE: 04-Mar-1991  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 89-09062  
; FILING DATE: 05-JUL-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meyers, Kenneth J.  
; REGISTRATION NUMBER: 25,146  
; REFERENCE/DOCKET NUMBER: 04958-0006-02000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-408-4000  
; TELEFAX: 202-408-4400  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 732 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-481-626-2

Query Match 14.0%; Score 62.5; DB 1; Length 732;  
Best Local Similarity 32.4%; Pred. No. 8.8;  
Matches 22; Conservative 8; Mismatches 19; Indels 19; Gaps 3;  
QY 2 DRRPLPVDRAAGLKEKTLILLDVSTKNPVRTV-NENFLSLQDPSIIH----- 48  
Db 154 ERAALP---AQELEBYNKILLDMETTYSVATVCHPNGSCQLQLEPDLTNVMATSRKYEDLL 210  
QY 49 ---DGWLD 53  
Db 211 WAWEGWRD 218  
RESULT 7  
US-08-989-299-4  
; Sequence 4, Application US/08989299  
; Patent No. 6194556  
; GENERAL INFORMATION:

APPLICANT: Acton, Susan L.  
APPLICANT: Robinson, Keith E.  
TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG  
TITLE OF INVENTION: AND THERAPEUTIC AND DIAGNOSTIC USES THEREFOR  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 11-DEC-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold E., Beth  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: MIA-025.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1306 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-989-299-7

Query Match 14.0%; Score 62.5; DB 4; Length 1306;  
Best Local Similarity 32.4%; Pred. No. 19;  
Matches 22; Conservative 8; Mismatches 19; Indels 19; Gaps 3;  
QY 2 DRRLPVDRAAGLKEKTLILLDVSTKNPVRTV-NENFLSLQDPSIIH----- 48  
Db 728 ERAALP---AQEEYVKILLDMETTVSVATVCHPNCGSCQLQPLDTNVMATSRKYEDLL 784  
QY 49 ----DGWLD 53  
Db 785 WAVEGWRD 792

RESULT 9  
US-08-751-474-2  
; Sequence 2, Application US/08751474  
; Patent No. 5821335  
; GENERAL INFORMATION:  
; APPLICANT: Hoskins, JoAnn  
; APPLICANT: Skatrud, Paul L.  
; TITLE OF INVENTION: Biosynthetic Gene murG From  
; TITLE OF INVENTION: Streptococcus Pneumoniae  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: U.S.  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Webster, Thomas D.  
; REGISTRATION NUMBER: 39, 872  
; REFERENCE/DOCKET NUMBER: X-9902  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3334  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 352 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-751-474-2

APPLICANT: Acton, Susan L.  
APPLICANT: Robinson, Keith E.  
TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG  
TITLE OF INVENTION: AND THERAPEUTIC AND DIAGNOSTIC USES THEREFOR  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 11-DEC-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold E., Beth  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: MIA-025.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 732 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-989-299-4

Query Match 14.0%; Score 62.5; DB 4; Length 732;  
Best Local Similarity 32.4%; Pred. No. 8.8; 19; Indels 19; Gaps 3;  
Matches 22; Conservative 8; Mismatches 19; Indels 19; Gaps 3;  
QY 2 DRRLPVDRAAGLKEKTLILLDVSTKNPVRTV-NENFLSLQDPSIIH----- 48  
Db 154 ERAALP---AQEEYVKILLDMETTVSVATVCHPNCGSCQLQPLDTNVMATSRKYEDLL 210  
QY 49 ----DGWLD 53  
Db 211 WAVEGWRD 218

RESULT 8  
US-08-989-299-7  
; Sequence 7, Application US/08989299  
; Patent No. 6194556  
; GENERAL INFORMATION:  
; APPLICANT: Acton, Susan L.  
; APPLICANT: Robinson, Keith E.  
; TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG  
; TITLE OF INVENTION: AND THERAPEUTIC AND DIAGNOSTIC USES THEREFOR  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```
;
;
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,857
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 251 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-483-857-9
;
;
; Query Match 13.2%; Score 59; DB 3; Length 251;
; Best Local Similarity 36.0%; Pred. No. 6;
; Matches 18; Conservative 7; Mismatches 21; Indels 4; Gaps 2;
;
; QY 4 RPLPVDRAAGLKERTLLLDVSTKNPVRTVNE-NFLSLQLDPSIIHDGWL 52
; DB 42 RAVEVD---GLSLATRTLVASTPGADWTPCOLDYTSRPLDVALQDDGWL 88
;
; RESULT 12
; US-08-698-407-4
; Sequence 4, Application US/08698407
; Patent No. 5856128
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN NUCLEIC ACID BINDING PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/698,407
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0116 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 475 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 267188
;
;
; Query Match 13.6%; Score 61; DB 2; Length 352;
; Best Local Similarity 32.3%; Pred. No. 5.2;
; Matches 20; Conservative 8; Mismatches 28; Indels 3; Gaps 3;
;
; QY 32 TVNENFSLQLDPSIIHDGW-LDFLSKRLV----TLARGLSPAFLEFG-GKRTDFLQFQ 85
; DB 11 TVGHVTLNLLMLPKFIEDGWEVHYIGDKGIEHQEILKSLGDLDTFHSIATGKLRREFSWQ 70
;
; QY 86 NL 87
; DB 71 NM 72
;
; RESULT 10
; US-09-188-930-305
; Sequence 305, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Orust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 305
; LENGTH: 649
; TYPE: PRT
; ORGANISM: Mouse
;
; US-09-188-930-305
;
; Query Match 13.6%; Score 61; DB 4; Length 649;
; Best Local Similarity 30.0%; Pred. No. 12;
; Matches 27; Conservative 11; Mismatches 34; Indels 18; Gaps 4;
;
; QY 5 PLPVDRAAGLKERTLLLDVSTKNPVR-----TVNENFSLQLDPSIIHDGWLDFL--SS 57
; DB 382 PAPVTQPDIKNEKLKIDQRTTGSPSRKTLITVK-----SVTPDTHISWRALPMTA 435
;
; QY 58 KRLVTLARGLSPAF-----LRFSGKRTDFL 82
; DB 436 LRLSLWKLGHSPAFSGSITETIVTGERSEYL 465
;
; RESULT 11
; US-08-483-857-9
; Sequence 9, Application US/08483857
; Patent No. 6020125
; GENERAL INFORMATION:
; APPLICANT: Chan, Voon Loong
; APPLICANT: Louie, Helena
; TITLE OF INVENTION: BASAL BODY ROD PROTEIN GENES OF
; TITLE OF INVENTION: CAMPYLOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
```





Search completed: July 30, 2002, 08:15:02  
Job time: 284 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:14:02 ; Search time 141.47 Seconds  
(without alignments)  
94.217 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_42\_161

Perfect score: 621

Sequence: 1 GDRPLPVDRAAGLKEKTLI.....YLLKNYEDDIVRSDVALDKQ 120

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	621	100.0	262	22	AA24147 Human EST encoded
2	621	100.0	534	22	AA24147 Heparanase-like pr
3	621	100.0	592	22	AA07424 Human heparanase-1
4	621	100.0	592	22	AA24147 Heparanase-like pr
5	621	100.0	592	22	AA24147 Human Heparanase-2
6	621	100.0	592	22	AA24147 Human Heparanase
7	615	99.0	492	22	AA24147 Amino acid sequenc
8	566	91.1	439	22	AA07423 Human heparanase-1
9	566	91.1	480	22	AA07418 Novel human extrac
10	566	91.1	480	22	AA24147 Heparanase-like pr
11	566	91.1	480	22	AA24147 Human heparanase,

12	565.5	91.1	538	22	AA24147 Human heparanase,
13	181.5	29.2	535	21	AA24147 A murine heparanas
14	156.5	25.2	532	20	AA24147 Seq ID No: 15 of W
15	156.5	25.2	543	20	AA24147 Human heparanase e
16	156.5	25.2	543	20	AA24147 A human heparanase
17	156.5	25.2	543	21	AA24147 Amino acid sequenc
18	156.5	25.2	543	21	AA24147 Human heparanase p
19	156.5	25.2	543	21	AA24147 Human heparanase
20	156.5	25.2	543	22	AA24147 Human heparanase i
21	156.5	25.2	543	22	AA24147 Human membrane or
22	156.5	25.2	543	22	AA24147 Human heparanase p
23	156.5	25.2	588	20	AA24147 A human protein wi
24	156.5	25.2	592	20	AA24147 A human heparanase
25	156.5	25.2	592	21	AA24147 Amino acid sequenc
26	155.5	25.0	530	20	AA24147 Human pre-prohepar
27	93	15.0	32	20	AA24147 Human pre-prohepar
28	86.5	13.9	488	22	AA24147 Amino acid sequenc
29	84.5	13.6	280	21	AA24147 Arabidopsis thalia
30	84.5	13.6	304	21	AA24147 Arabidopsis thalia
31	84.5	13.6	304	21	AA24147 Arabidopsis thalia
32	82.5	13.3	280	21	AA24147 Arabidopsis thalia
33	82.5	13.3	302	21	AA24147 Arabidopsis thalia
34	82.5	13.3	304	21	AA24147 Arabidopsis thalia
35	82.5	13.3	488	22	AA24147 Amino acid sequenc
36	80.5	13.0	488	22	AA24147 Amino acid sequenc
37	76.5	12.3	488	22	AA24147 Drosophila melanog
38	75	12.1	241	22	AA24147 Novel human diagno
39	74	11.9	402	22	AA24147 Human polypeptide
40	73.5	11.8	445	22	AA24147 Human polypeptide
41	73.5	11.8	445	22	AA24147 Human polypeptide
42	73.5	11.8	445	22	AA24147 Human polypeptide
43	73.5	11.8	641	22	AA24147 Human polypeptide
44	73.5	11.8	698	22	AA24147 Human polypeptide
45	73.5	11.8	698	22	AA24147 Human polypeptide

ALIGNMENTS

RESULT	1
AA24147	AA24147
ID	AA24147 standard; Protein; 262 AA.
XX	AA24147;
AC	AC
XX	XX
DT	12-OCT-2001 (first entry)
XX	XX
DE	Human EST encoded protein SEQ ID NO: 1672.
XX	XX
KW	Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW	diagnostics; forensic test; gene mapping; genetic disorder;
KW	biodiversity; gene therapy; nutrition.
XX	XX
OS	Homo sapiens.
XX	XX
PN	WO200154477-A2.
XX	XX
PD	02-AUG-2001.
XX	XX
PF	25-JAN-2001; 2001WO-US02687.
XX	XX
PR	25-JAN-2000; 2000US-0491404.
PR	17-JUL-2000; 2000US-0617746.
PR	03-AUG-2000; 2000US-0631451.
PR	15-SEP-2000; 2000US-0663870.
XX	XX
PA	(HYSE-) HYSEQ INC.
XX	XX
PI	Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI	Cao Y, Drmanac RA, Zhang J, Werhman T;
XX	XX
DR	WPI; 2001-476164/51.

DR N-PSDB; AAH98806.  
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
 PT antibodies and research use -  
 PT  
 XX  
 PS Claim 20; Page 1122-1123; 1275pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences of novel  
 CC proteins from a variety of organisms, including human, dog, cat, horse,  
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 CC from the organism of interest. They can be used in diagnostics,  
 CC forensics, gene mapping, identification of mutations, to assess  
 CC biodiversity and for nutritional purposes. The present sequence is a  
 CC protein of the invention.  
 XX  
 SQ Sequence 262 AA;  
 Query Match 100.0%; Score 621; DB 22; Length 262;  
 Best Local Similarity 100.0%; Pred. No. 9.1e-65;  
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GDRRLPVDRAAGLKEKTLILLIDVSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSSKRL 60  
 DB 42 GDRRLPVDRAAGLKEKTLILLIDVSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSSKRL 101  
 QY 61 VTLAGLSPAFRLFGGKRTDFLQFONLRNPAKSRGGPGPDYLYKNYEDDIVRSDVALDKQ 120  
 DB 102 VTLAGLSPAFRLFGGKRTDFLQFONLRNPAKSRGGPGPDYLYKNYEDDIVRSDVALDKQ 161  
 RESULT 2  
 AAB85216  
 ID AAB85216 standard; Protein; 534 AA.  
 XX  
 AC AAB85216;  
 XX  
 DT 07-SEP-2001 (first entry)  
 XX  
 DE Heparanase-like protein Hpa2 splice variant #2.  
 XX  
 KW Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;  
 KW cytosolic; neuroprotective; cerebroprotective; immunosuppressive;  
 KW antiproliferative; neurotropic; antiinflammatory; antiarthritic; antiasthmatic;  
 KW antidiabetic; antiarteriosclerotic; vulnerary.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200146392-A2.  
 XX  
 PD 28-JUN-2001.  
 XX  
 PF 21-DEC-2000; 2000WO-GB04963.  
 XX  
 PR 22-DEC-1999; 99GB-0030392.  
 XX  
 PR 07-APR-2000; 2000GB-0008713.  
 XX  
 PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
 XX  
 XX McKenzie EA, Stamps AC, Terrett JA, Tyson KL;  
 PI  
 XX WPI: 2001-418056/44.  
 DR  
 DR N-PSDB; AAH22672.  
 XX  
 XX Novel homologs of heparanase, present in three splice variants, useful  
 PT for identifying agents that modulate heparanase, useful in the  
 PT treatment and/or prophylaxis of abnormal levels of heparanase -  
 XX  
 PS Claim 1; Fig 2; 97pp; English.  
 XX  
 CC The invention provides a homologue to heparanase which is present in  
 CC three splice variants. The heparanase homologue polypeptide is useful in

CC the treatment of a human or non-human animal or for use in diagnosis.  
 CC Vectors comprising the heparanase homologue polynucleotides are useful in  
 CC the transformation or transfection of a prokaryotic or eukaryotic host.  
 CC The modulators of the polypeptide are useful in the manufacture of a  
 CC medicament for the treatment and/or prophylaxis of a condition/disease  
 CC associated with abnormal levels of the heparanase homologue, including  
 CC cancer, central nervous system (CNS) and neurodegenerative diseases,  
 CC cardiovascular diseases such as stenosis following angioplasty and  
 CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,  
 CC allografts, inflammatory diseases, arthritis, vascular restenosis,  
 CC tumour growth and progression, asthma, Alzheimer's disease, diabetic  
 CC retinopathy, wound healing and inflammation. The polypeptide is also  
 CC useful in diagnosis and research. The present sequence represents the  
 CC amino acid sequence of the mid-sized splice variant of the heparanase-  
 CC like protein Hpa2 of the invention.  
 XX  
 SQ Sequence 534 AA;  
 Query Match 100.0%; Score 621; DB 22; Length 534;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-64;  
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GDRRLPVDRAAGLKEKTLILLIDVSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSSKRL 60  
 DB 42 GDRRLPVDRAAGLKEKTLILLIDVSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSSKRL 101  
 QY 61 VTLAGLSPAFRLFGGKRTDFLQFONLRNPAKSRGGPGPDYLYKNYEDDIVRSDVALDKQ 120  
 DB 102 VTLAGLSPAFRLFGGKRTDFLQFONLRNPAKSRGGPGPDYLYKNYEDDIVRSDVALDKQ 161  
 RESULT 3  
 AAU07424  
 ID AAU07424 standard; Protein; 592 AA.  
 XX  
 AC AAU07424;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Human heparanase-like protein splice variant #1.  
 XX  
 KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
 KW antiproliferative; cardiant; vasotropic; cerebroprotective; neurotropic;  
 KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
 KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;  
 KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
 KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; sunburn;  
 KW nervous system disorder; Alzheimer's disease; ocular disorder; infection;  
 KW wound healing; food additive; heparanase.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200179253-A1.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 11-APR-2001; 2001WO-US11643.  
 XX  
 PR 18-APR-2000; 2000US-198123P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Fiscella M, Shi Y, Ebner R, Ruben SM;  
 PI  
 XX WPI: 2001-611720/70.  
 DR  
 DR N-PSDB; AAS13848.  
 XX  
 XX New nucleic acids encoding extracellular matrix polypeptides, for  
 PT diagnosing, treating, preventing or ameliorating human disorders and  
 PT disease, such as, autoimmune, hyperproliferative or cardiovascular  
 PT disorders -  
 XX



XX PF 11-SEP-2000; 2000WO-EP08837.  
 XX PR 23-SEP-1999; 99EP-0118805.  
 XX PR 07-JUL-2000; 2000EP-0114649.  
 XX PA (MERE ) MERCK PATENT GMBH.  
 XX PI Duecker K, Sirrenberg C;  
 XX DR WPI; 2001-308089/32.  
 XX DR N-PSDB; AAF86101.  
 XX PT New heparanase-2 polypeptide useful in diagnosing (the susceptibility  
 of a subject to) and as vaccines against e.g. autoimmune disorders,  
 PT cardiovascular disease, cancer, diabetes, ischemia, sepsis, stroke, or  
 PT thrombosis -  
 XX PS Claim 1; Page 42-43; 45pp; English.  
 XX CC This invention relates to a human heparanase-2 protein and the cDNA  
 sequence encoding it. Heparanase-2 is a member of the endoglucuronidase  
 family of polypeptides and it degrades heparan sulphate proteoglycans  
 CC HSPGs (ubiquitous macromolecules of cell surfaces, basement membranes and  
 CC the extracellular matrix). HSPGs support the vascular endothelium and  
 CC stabilise the structure of the capillary wall. Heparanases may be  
 CC associated with neovascularization and metastasis related to malignant  
 CC tumour formation. Heparanase-2 polynucleotides and proteins are useful as  
 CC vaccines for inducing an immunological response against autoimmune  
 CC disorders, blood coagulation disorders, cancer, diabetes, ischaemia,  
 CC sepsis, stroke, cardiovascular diseases, or thrombosis, as well as in  
 CC diagnosing (the susceptibility of a subject to) these diseases.  
 CC Heparanase-2 fragments may be used as immunogens to produce antibodies  
 CC immunospecific to the polypeptides, and to identify membrane bound  
 CC soluble receptors, agonists or antagonists that compete with the binding  
 CC of the polypeptide to the receptors. An antibody specific for  
 CC heparanase-2 can be used in the diagnosis of the above diseases and in  
 CC isolating or identifying clones expressing heparanase-2. The present  
 CC sequence represents heparanase-2. Three regions of heparanase-2 with high  
 CC immunogenicity (immunogenic epitopes) can be used to raise antibodies  
 CC against heparanase-2.  
 XX SQ Sequence 592 AA;

Query Match 100.0%; Score 621; DB 22; Length 592;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-64;  
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQIDPSIHDGWLDFLSSKRL 60  
 DB 42 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQIDPSIHDGWLDFLSSKRL 101  
 QY 61 VTLAGLSPAFIRFGKRTDFIQNLNPAKSRGGPGDYLYKNYEDDIVRSDVALDKQ 120  
 DB 102 VTLAGLSPAFIRFGKRTDFIQNLNPAKSRGGPGDYLYKNYEDDIVRSDVALDKQ 161

RESULT 6  
 AAY97632  
 ID AAY97632 standard; Protein; 592 AA.  
 XX AC AAY97632;  
 XX DT 20-APR-2001 (first entry)  
 XX DE Human heparanase, hnhp1, protein sequence.  
 XX KW Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;  
 KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;  
 KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;  
 KW gene therapy; human.  
 XX PF 18-DEC-2000; 2000WO-EP12909.

OS Homo sapiens.  
 XX PN WO200100643-A2.  
 XX XX 04-JAN-2001.  
 XX PF 19-JUN-2000; 2000WO-IL00358.  
 XX PR 25-JUN-1999; 99US-0140801.  
 XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 XX PI Pecker I, Michal I, Itzhaki H;  
 XX DR WPI; 2001-137930/14.  
 XX DR N-PSDB; AAA91097.  
 XX PT New polynucleotides and polypeptides that are distantly homologous to  
 PT heparanase, useful in wound healing, as well as in gene therapy  
 PT protocols for angiogenesis, restenosis, atherosclerosis, or  
 PT inflammation -  
 XX PS Claim 10; Fig 1; 67pp; English.  
 XX CC This sequence represents a heparanase of the invention.  
 CC The heparanase DNA and protein sequences are useful in wound healing,  
 CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary  
 CC diseases, neurodegenerative diseases (such as Scrape, Alzheimer's  
 CC disease, and Creutzfeldt-Jakob disease) or viral infections. The  
 CC heparanase coding sequence is particularly useful in gene therapy.  
 XX SQ Sequence 592 AA;

Query Match 100.0%; Score 621; DB 22; Length 592;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-64;  
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQIDPSIHDGWLDFLSSKRL 60  
 DB 42 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQIDPSIHDGWLDFLSSKRL 101  
 QY 61 VTLAGLSPAFIRFGKRTDFIQNLNPAKSRGGPGDYLYKNYEDDIVRSDVALDKQ 120  
 DB 102 VTLAGLSPAFIRFGKRTDFIQNLNPAKSRGGPGDYLYKNYEDDIVRSDVALDKQ 161

RESULT 7  
 AAB84664  
 ID AAB84664 standard; Protein; 492 AA.  
 XX AC AAB84664;  
 XX DT 05-SEP-2001 (first entry)  
 XX DE Amino acid sequence of human heparanase-like polypeptide.  
 XX KW Human; heparanase-like polypeptide; gene therapy; cancer; angiogenesis;  
 KW trauma; autoimmune disease; skin disease; cardiovascular disease;  
 KW nervous system disease; inflammation; arthritis; genitalia;  
 KW male fertility; erectile dysfunction.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 FT Misc-difference 407 /note= "unspecified residue encoded by KCA"  
 XX PN WO200148161-A2.  
 XX PD 05-JUL-2001.  
 XX PF 18-DEC-2000; 2000WO-EP12909.

```
XX PR 23-DEC-1999; 99EP-0125831.
XX XX (SCHD ) SCHERING AG.
XX PA
XX XX Siemeister G, Weiss B;
XX PI
XX XX WPI; 2001-418259/44.
XX DR N-PSDB; AAH28347.
XX XX
XX XX Human Heparanase-like polynucleotide encoding polypeptides useful for
PT modulating expression of the polypeptide and for treating cancer,
PT cancer metastasis, aberrant angiogenesis by gene therapy technique -
XX XX
XX PS Claim 9; Page 30; 30pp; English.
XX XX
XX CC The present sequence represents a human heparanase-like polypeptide.
XX CC Heparanase-like polynucleotides are useful as a source of probes,
XX CC primers and antisense molecules, and in gene therapy. Heparanase-like
XX CC polynucleotides and polypeptides are useful for treating several
XX CC disorders e.g., cancer, cancer metastasis. The oligonucleotides are
XX CC also useful as diagnostic markers for the diagnosis of disorder such
XX CC as cancer, cancer metastasis and aberrant angiogenesis. They may also
XX CC act as diagnostic markers for diagnosis of disorder such as cancer,
XX CC cancer metastasis and aberrant angiogenesis. The heparanase polypeptides
XX CC and polynucleotides are also useful for treating trauma, autoimmune
XX CC diseases, skin diseases, cardiovascular diseases, nervous system
XX CC diseases, and inflammation including arthritis. Since the polynucleotide
XX CC is preferentially expressed in male genitalia, modulation of its
XX CC expression and/or activity may be used for medical intervention in male
XX CC genitalia function that is male fertility control, erectile dysfunction.
XX XX
XX SQ Sequence 492 AA;

Query Match 99.08; Score 615; DB 22; Length 492;
Best Local Similarity 100.0%; Pred. No. 1.1e-63;
Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DRRPLPVDRRAAGLKEKTLILLDVSTKNPVRTVNFNLSLQLDPSIIHDGWLDFLSSKRLV 61
DB 1 drrplpvdraaglkectlllldvstknpvrtnvnenflslqldpslihdgwlfdlsskrlv 60

QY 62 TLARGLSPAFURFGGKRTDFLQFQNLNRPKASRGPGPDYLYLKNYEDDIVRSDVALDKQ 120
DB 61 tlarglspafirfggkrtdfiqfqlnlnrpaksgpgpdylylknayeddivrsdvaldkq 119

RESULT 8
AAU07423
ID AAU07423 standard; Protein; 439 AA.
XX AC
XX AC AAU07423;
XX DT 18-DEC-2001 (first entry)
XX DE
XX DE Human heparanase-like protein.
XX KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW antiproliferative; cardiant; vasotropic; cerebroprotective; neurotropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;
KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;
XX KW wound healing; food additive; heparanase.
XX OS Homo sapiens.
XX XX
XX PN WO200179253-A1.
XX XX
XX PD 25-OCT-2001.
XX XX
```

```
PF 11-APR-2001; 2001WO-US11643.
XX PR
XX PR 18-APR-2000; 2000US-198123P.
XX PA
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI
XX PI Fiscella M, Shi Y, Ebner R, Ruben SM;
XX XX WPI; 2001-611720/70.
XX DR
XX XX
XX XX New nucleic acids encoding extracellular matrix polypeptides, for
PT diagnosing, treating, preventing or ameliorating human disorders and
PT disease, such as, autoimmune, hyperproliferative or cardiovascular
PT disorders -
XX XX
XX PS Disclosure; Page 13-14; 308pp; English.
XX XX
XX CC The invention relates to novel isolated polynucleotides (I) encoding
XX CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by
XX CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.
XX CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They
XX CC are also used in diagnosing a pathological condition or susceptibility to
XX CC a pathological condition. The antibodies to the polypeptides can also be
XX CC used in alleviating symptoms associated with the disorders and in
XX CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
XX CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
XX CC include autoimmune diseases e.g. rheumatoid arthritis,
XX CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
XX CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
XX CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
XX CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
XX CC and ocular disorders e.g. corneal infection. The polypeptides can also
XX CC be used to aid wound healing and epithelial cell proliferation, to
XX CC prevent skin aging due to sunburn, to maintain organs before
XX CC transplantation, for supporting cell culture of primary tissues, to
XX CC regenerate tissues and in chemotaxis. The polypeptides can also be used
XX CC as a food additive or preservative to increase or decrease storage
XX CC capabilities. The present sequence represents the amino acid sequence
XX CC of human heparanase-like protein.
XX XX
XX SQ Sequence 439 AA;

Query Match 91.1%; Score 566; DB 22; Length 439;
Best Local Similarity 99.1%; Pred. No. 5.4e-58;
Matches 108; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVDRRAAGLKEKTLILLDVSTKNPVRTVNFNLSLQLDPSIIHDGWLDFLSSKRL 60
DB 1 gdrplpvdraaglkectlllldvstknpvrtnvnenflslqldpslihdgwlfdlsskrl 60

QY 61 VTLARGLSPAFURFGGKRTDFLQFQNLNRPKASRGPGPDYLYLKNYEDD 109
DB 61 vtlarglspafirfggkrtdfiqfqlnlnrpaksgpgpdylylknayed 109

RESULT 9
AAU07418
ID AAU07418 standard; Protein; 480 AA.
XX AC
XX AC AAU07418;
XX DT 18-DEC-2001 (first entry)
XX DE
XX DE Novel human extracellular matrix (ECM) protein #1.
XX KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW antiproliferative; cardiant; vasotropic; cerebroprotective; neurotropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;
KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;
```





DE Human heparanase, hnhp1 pn5 form, protein sequence.

KW Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;  
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;  
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;  
KW gene therapy; human.

OS Homo sapiens.

XX WO200100643-A2.

XX 04-JAN-2001.

XX 19-JUN-2000; 2000WO-IL00358.

XX 25-JUN-1999; 99US-0140801.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX Pecker I, Michal I, Itzhaki H;

XX WPI; 2001-137930/14.

XX N-PSDB; AAA91099.

XX New polynucleotides and polypeptides that are distantly homologous to  
PT heparanase, useful in wound healing, as well as in gene therapy  
PT protocols for angiogenesis, restenosis, atherosclerosis, or  
PT inflammation -

XX Claim 10; Page 63; 67pp; English.

XX This sequence represents a heparanase of the invention.

CC The heparanase DNA and protein sequences are useful in wound healing,  
CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary  
CC diseases, neurodegenerative diseases (such as Scrape, Alzheimer's  
CC disease, and Creutzfeldt-Jakob disease) or viral infections. The  
CC heparanase coding sequence is particularly useful in gene therapy.

XX SQ Sequence 480 AA;

Query Match 91.1%; Score 566; DB 22; Length 480;

Best Local Similarity 99.1%; Pred. No. 6.1e-58;

Matches 108; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GDRRPLPVDRAGLKEKTLILLDYSTKNPVTNENFLSLQDPSIIHDGWLDFLSSKRL 60

Db 42 GDRPLPVDRAGLKEKTLILLDYSTKNPVTNENFLSLQDPSIIHDGWLDFLSSKRL 101

Qy 61 VTLARGSPAPLRGGRKTFDQFQNLNPNPAKSRGGPGDYLYLKNYEDD 109

Db 102 vtlargispaflrfggrktdflqnlrnpaksggpgpdylylknede 150

RESULT 12

AA97633

ID AA97633 standard; Protein: 538 AA.

XX AC

XX AA97633;

DT 20-APR-2001 (first entry)

XX Human heparanase, hnhp1 pn9 form, protein sequence.

XX Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;  
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;  
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;  
KW gene therapy; human.

OS Homo sapiens.

XX XX

XX Key Location/Qualifiers

XX Misc-difference 305

FT

PT /note= "encoded by GAC"

XX WO200100643-A2.

XX 04-JAN-2001.

XX 19-JUN-2000; 2000WO-IL00358.

XX 25-JUN-1999; 99US-0140801.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX Pecker I, Michal I, Itzhaki H;

XX WPI; 2001-137930/14.

XX N-PSDB; AAA91099.

XX New polynucleotides and polypeptides that are distantly homologous to  
PT heparanase, useful in wound healing, as well as in gene therapy  
PT protocols for angiogenesis, restenosis, atherosclerosis, or  
PT inflammation -

XX Claim 10; Page 61-62; 67pp; English.

XX This sequence represents a heparanase of the invention.

CC The heparanase DNA and protein sequences are useful in wound healing,  
CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary  
CC diseases, neurodegenerative diseases (such as Scrape, Alzheimer's  
CC disease, and Creutzfeldt-Jakob disease) or viral infections. The  
CC heparanase coding sequence is particularly useful in gene therapy.

XX SQ Sequence 538 AA;

Query Match 91.1%; Score 565.5; DB 22; Length 538;

Best Local Similarity 95.0%; Pred. No. 8.1e-58;

Matches 113; Conservative 0; Mismatches 1; Indels 5; Gaps 2;

Qy 1 GDRRPLPVDRAGLKEKTLILLDYSTKNPVTNENFLSLQDPSIIHDGWLDFLSSKRL 60

Db 42 GDRPLPVDRAGLKEKTLILLDYSTKNPVTNENFLSLQDPSIIHDGWLDFLSSKRL 101

Qy 61 VTLARGSPAPLRGGRKTFDQFQNLNPNPAKSRGGPGDYLYLKNYEDD 119

Db 102 vtlargispaflrfggrktdflqnlrnpaksggpgpdylylknedy--ars---ldk 155

RESULT 13

AA08851

ID AA08851 standard; Protein: 535 AA.

XX AC

XX AA08851;

DT 15-JAN-2001 (first entry)

XX A murine heparanase polypeptide.

XX Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;  
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;  
KW wound healing; infection; burn; angiogenesis; restenosis;  
KW atherosclerosis; inflammation; neurodegenerative disease;  
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.

XX Mus sp.

XX WO200052178-A1.

XX 08-SEP-2000.

XX 14-FEB-2000; 2000WO-US03542.

XX 01-MAR-1999; 99US-0258892.

XX XX

```

PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
XX Pecker I, Vlodavsky I, Feinstein E;
XX
XX WPI: 2000-579289/54.
XX N-PSDB; AAX75081.
XX
XX New polynucleotides encoding a polypeptide having heparanase activity,
XX useful in wound healing and in gene therapy, particularly in treating
XX tumour, inflammation, autoimmunity, neurodegenerative diseases
XX
XX Claim 22: Page 144-145; 152pp; English.
XX
XX The present sequence represents murine protein with heparanase catalytic
XX activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
XX particularly in treating tumour, inflammation or autoimmunity.
XX Particularly, the polynucleotide is useful in modulating the
XX bioavailability of heparin-binding growth factors, cellular responses
XX to heparin-binding growth factors (e.g. bFGF) and cytokines
XX (e.g. interleukin (IL)-8), cell interaction with plasma lipoproteins,
XX cellular susceptibility to certain viral and some bacterial and protozoa
XX infections, or disintegration of neurodegenerative plaques. The
XX polynucleotide is also useful in wound healing (e.g. thermal, chemical
XX or radiation burns), and in the treatment of angiogenesis, restenosis,
XX atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
XX Straussler Syndrome or Creutzfeldt-Jakob disease), and some viral,
XX bacterial or protozoa infections.
XX
XX Sequence 535 AA;

Query Match 29.2%; Score 181.5; DB 21; Length 535;
Best Local Similarity 45.7%; Pred. No. 1.3e-12;
Matches 43; Conservative 14; Mismatches 32; Indels 5; Gaps 2;

QY 22 LDVSTKNPVRTVNENFLSLQDPSIIHD-GWLDLFLSKRLVTLARGLSPAFLRFGGKRTD 80
| | | | | : | | | : | | | | | : | | | | | : | | | | | : | | | | | :
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 33 leftytkprlsvspfsitidaalatdprftlfgsprlralarglspaylrfggktkd 92

QY 81 FLQFQNLNPAKSRGGPGDYLLKNEYDDIVRSD 114
| | | | | : | | | : | | | | | : | | | | | : | | | | | : | | | | | :
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 93 flifl----dpdkeptseersywksgvnhdicrse 122

RESULT 14
AAY17083
XX
XX AAY17083 standard; Protein: 532 AA.
XX
XX AAY17083;
XX
XX 21-JUL-1999 (first entry)
XX
XX Seq ID No: 15 of WO9921975.
XX
XX Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
XX metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
XX arteriosclerosis; atherosclerosis; inflammation; tissue development;
XX human; HSPG.
XX
XX Homo sapiens.
XX
XX WO9921975-A1.
XX
XX 06-MAY-1999.
XX
XX 28-OCT-1998; 98WO-AU00898.
XX
XX 09-DEC-1997; 97AU-0000812.
XX
XX 28-OCT-1997; 97AU-0000062.
XX
XX (AUSU ) UNIV AUSTRALIAN NAT.
XX
XX Freeman CG, Hamdorf BJ, Hulett MD, Parish CR;
XX
XX WPI; 1999-312956/26.
XX
XX N-PSDB; AAX37259.

PA
XX
XX Freeman CG, Hamdorf BJ, Hulett MD, Parish CR;
XX
XX WPI: 1999-312956/26.
XX
XX N-PSDB; AAX37260.
XX
XX Polynucleotides encoding mammalian endoglucuronidases, especially
XX heparanases, useful to promote wound healing
XX
XX Claim 6: Page 76-79; 112pp; English.
XX
XX The invention relates to nucleic acid sequences that encode heparanase
XX enzymes having endoglucuronidase activity. Recombinant heparanases are
XX capable of removing the HS side chain from heparan sulfate proteoglycan
XX (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
XX inhibit heparanase, this is useful for treatment of a physiological or
XX medical condition associated with elevated heparanase activity, such as
XX metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
XX arteriosclerosis, atherosclerosis and inflammation. The human, murine and
XX rat heparanases can be used to enhance wound healing, especially
XX associated with tissue development and repair. The conditions mentioned
XX above can be diagnosed using specific antibodies, and also using primers
XX and probes specific for the heparanase polynucleotides. Other uses of the
XX heparanases include sequencing sulfated molecules such as HSPG.
XX
XX Sequence 532 AA;

Query Match 25.2%; Score 156.5; DB 20; Length 532;
Best Local Similarity 37.2%; Pred. No. 1.1e-09;
Matches 42; Conservative 20; Mismatches 44; Indels 7; Gaps 3;

QY 1 GDRRPLPVDRAAGLKETLILLDVSTKNPVRTVNENFLSLQDPSIIHD-GWLDLFLSKR 59
| | | | | : | | | : | | | | | : | | | | | : | | | | | : | | | | | :
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 22 gpispgalprpa--qaqdvvdldfftqephlvpsflsvtdanlatdprflllgspk 79

QY 60 LVTLARGLSPAFLRFGGKRTDFLQFQNLNPAKSRGGPGDYLLKNEYDDIVR 112
| | | | | : | | | : | | | | | : | | | | | : | | | | | : | | | | | :
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 80 lrtlarglspaylrfggktkdflif----dpkkestfeersywgsgvqnqcdick 128

RESULT 15
AAY17082
XX
XX AAY17082 standard; Protein: 543 AA.
XX
XX AAY17082;
XX
XX 21-JUL-1999 (first entry)
XX
XX Human heparanase enzyme.
XX
XX Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
XX metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
XX arteriosclerosis; atherosclerosis; inflammation; tissue development;
XX human; HSPG.
XX
XX Homo sapiens.
XX
XX WO9921975-A1.
XX
XX 06-MAY-1999.
XX
XX 28-OCT-1998; 98WO-AU00898.
XX
XX 09-DEC-1997; 97AU-0000812.
XX
XX 28-OCT-1997; 97AU-0000062.
XX
XX (AUSU ) UNIV AUSTRALIAN NAT.
XX
XX Freeman CG, Hamdorf BJ, Hulett MD, Parish CR;
XX
XX WPI; 1999-312956/26.
XX
XX N-PSDB; AAX37259.

```

XX Polynucleotides encoding mammalian endoglucuronidases, especially  
PT heparanases, useful to promote wound healing  
XX  
XX  
PS Claim 6; Page 69-73; 112pp; English.  
XX  
XX The invention relates to nucleic acid sequences that encode heparanase  
CC enzymes having endoglucuronidase activity. Recombinant heparanases are  
CC capable of removing the HS side chain from heparan sulfate proteoglycan  
CC (HSPG). Sulfated oligosaccharides, sulphates or HSPG can be used to  
CC inhibit heparanase, this is useful for treatment of a physiological or  
CC medical condition associated with elevated heparanase activity, such as  
CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,  
CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and  
CC rat heparanases can be used to enhance wound healing, especially  
CC associated with tissue development and repair. The conditions mentioned  
CC above can be diagnosed using specific antibodies, and also using primers  
CC and probes specific for the heparanase polynucleotides. Other uses of the  
CC heparanases include sequencing sulfated molecules such as HSPG. The  
CC present sequence represents a human heparanase.  
XX  
SQ Sequence 543 AA;

Query Match 25.2%; Score 156.5; DB 20; Length 543;  
Best Local Similarity 37.2%; Pred. No. 1.2e-09;  
Matches 42; Conservative 20; Mismatches 44; Indels 7; Gaps 3;  
QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLOLDPSIIHD-GWLDFLSSKR 59  
Db | | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
22 gplspgalprpa--gaqdvvdldfftqepihlvspfsivtdanlatdprfllllgspk 79  
QY 60 LVTTLARGLSPAFLRFGKRTDFLOFQNLRNPAKSRGPGPDYLYLKNYEDDIVR 112  
Db | | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
80 lrtlarginpaylrfggktgdlif----dpkkestfeersywsqsvngdqick 128

Search completed: July 30, 2002, 08:14:02  
Job time: 339 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:16:27 ; Search time 71.68 Seconds  
(without alignments)  
160.864 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_42\_161  
Perfect score: 621  
Sequence: 1 GDRRLPVDRAAGLKEKTLI.....YLLKNYEDIVRSVALDKQ 120

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	566	91.1	480	JC7506	heparanase protein
2	85.5	13.8	1391	T20642	hypothetical prote
3	85.5	13.8	1397	F87998	protein F09C3.1 [i
4	84.5	13.6	334	A86397	protein T7N9.6 [im
5	77.5	12.5	481	S56299	hypothetical prote
6	76	12.2	332	A81675	conserved hypothet
7	75	12.1	329	E71516	hypothetical prote
8	74	11.9	150	T34697	probable insertion
9	73.5	11.8	253	AF3625	cell division inh
10	73.5	11.8	581	T12095	beta-fructofuranos
11	72	11.6	160	T37000	probable transpos
12	72	11.6	293	T09171	ribosomal protein
13	72	11.6	293	T09170	ribosomal protein
14	72	11.6	582	A84206	hypothetical prote
15	71.5	11.5	1283	S52500	SNH1 protein homol
16	70.5	11.4	188	AC3613	probable s-adenosy
17	70	11.3	1234	S52099	phospholipase C be
18	70	11.3	1234	T38994	phospholipase C-be
19	70	11.3	1983	T00385	KIAA0624 protein -
20	69	11.1	378	AC3523	aliphatic sulfonat
21	68.5	11.0	378	A81275	hypothetical prote
22	68.5	11.0	613	JC6033	mosquitocidal prot
23	68.5	11.0	694	F90609	exodeoxyribonuclea
24	68	11.0	900	C64232	alanine--trNA liga
25	68	11.0	1084	B64088	hemoglobin-binding
26	67.5	10.9	165	T36981	probable insertion
27	67.5	10.9	492	S71146	H+-transporting AT
28	67	10.8	599	S18735	centromere protein
29	67	10.8	742	S58691	kinesin-related pr

30 66 10.6 289 2 A12165 bicarbonate transp  
31 66 10.6 367 2 JC2560 flagellin fliC-2 -  
32 65.5 10.5 332 2 C95150 SAP domain protein  
33 65.5 10.5 433 2 C82899 thymidine phosphor  
34 65 10.5 327 2 T39981 rev interacting pr  
35 65 10.5 652 2 D96808 protein F28K19.6 [i  
36 65 10.5 830 2 F83659 stage II sporulati  
37 65 10.5 1299 2 AH2090 two-component hybr  
38 65 10.5 2288 2 T30568 acetyl-CoA carboxy  
39 64.5 10.4 251 1 XMEBPF flagellar basal bo  
40 64.5 10.4 251 2 AC0640 probable flagellar  
41 64.5 10.4 296 2 A64035 hypothetical prote  
42 64 10.3 274 2 G96799 hypothetical prote  
43 64 10.3 332 2 S77386 nitrate transport  
44 64 10.3 345 2 S59236 SAE2 protein - yea  
45 64 10.3 363 2 S44443 RAD23 protein homo

ALIGNMENTS

RESULT 1  
JC7506  
heparanase protein 2a - human  
C:Species: Homo sapiens (man)  
C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 01-Dec-2000  
C:Accession: JC7506  
R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hirccock, M.;  
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000  
A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase fami  
A:Reference number: JC7506  
A:Accession: JC7506  
A:Molecule type: mRNA  
A:Residues: 1-480 <MCK>  
A:Cross-references: GB:AF282885  
C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and th  
therapies.  
C:Genetics:  
A:Gene: hpa2a  
A:Map position: 10q23-10q24  
C:Keywords: heparin binding; membrane bound

Query Match 91.1%; Score 566; DB 2; Length 480;  
Best Local Similarity 99.1%; Pred. No. 5.8e-53;  
Matches 108; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 GDRRLPVDRAAGLKEKTLIILLDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSKRL 60  
Db 42 GDRRLPVDRAAGLKEKTLIILLDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSKRL 101  
Oy 61 VTLARGLSPAFIRFGGKRTDFLQFONLRNPAKSRGGPGPDYLLKNYEDD 109  
Db 102 VTLARGLSPAFIRFGGKRTDFLQFONLRNPAKSRGGPGPDYLLKNYED 150

RESULT 2  
T20642  
hypothetical protein F09C3.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T20642  
R:Lennard, N.  
submitted to the EMBL Data Library, March 1997  
A:Reference number: Z19304  
A:Accession: T20642  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1391 <NIL>  
A:Cross-references: EMBL:Z92781; PIDN:CA807179.2; GSPDB:GN00019; CESP:F09C3.1  
A:Experimental source: clone F09C3  
C:Genetics:  
A:Gene: CESP:F09C3.1



Db 334 KTVIPAKVFGKFSIRTVPMDSSEKUTSLVQKHCDKAFKSLNSPNKCRTELHIDGAYWYSD 393  
Qy 53 ----DFLSSKRLVTLARGLSPAFLRFGGKRTDFLQFQNLNPA-----KSRGGPG 98  
Db 394 PFNAQFTAAKATKLVGVDPDFTRFGSGIPITLTFQDALNTSVLLPWRGDDG 448  
RESULT 6  
A81675  
conserved hypothetical protein TC0713 [imported] - Chlamydia muridarum (strain Nigg)  
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C:Accession: A81675  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: A81500; MUID:20150255  
A:Accession: A81675  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-332 <TET>  
A:Cross-references: GB:AE002339; GB:AE002160; NID:g7190728; PIDN:AAF39526.1; PID:g719074  
A:Experimental source: strain Nigg (MoPn)  
C:Genetics:  
A:Gene: TC0713

Query Match 12.2%; Score 76; DB 2; Length 332;  
Best Local Similarity 28.8%; Pred. No. 2.3;  
Matches 34; Conservative 19; Mismatches 45; Indels 20; Gaps 6;  
Qy 3 RRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSKRLVT 62  
Db 202 KGRPPKQSAKVEAETVSSDIYTKVP--SAARRELF--PEITSASLTF--SEKFT 254  
Qy 63 LARGLSPAFLRFGGKRTDFL-----QFQNLNPAKSRG-----GPGPDYILKNYEDD 109  
Db 255 EEEFL--AHLRGGGRLEQNLAKFSERFDSLRELSAKILGYDGDGDSDFGEEYDD 310

RESULT 7  
E71516  
hypothetical protein CT429 - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
C:Accession: E71516  
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra  
A:Reference number: A71570; MUID:99000809  
A:Accession: E71516  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-329 <ARN>  
A:Cross-references: GB:AE001316; GB:AE001273; NID:g3328850; PIDN:AAC68026.1; PID:g332885  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics:  
A:Gene: CT429

Query Match 12.1%; Score 75; DB 2; Length 329;  
Best Local Similarity 29.7%; Pred. No. 2.9;  
Matches 35; Conservative 19; Mismatches 44; Indels 20; Gaps 6;  
Qy 3 RRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSKRLVT 62  
Db 202 KGRPPKQSAKVEAETVSSDIYTKVP--SVARRELF--PEITSASLTF--SEKFT 254  
Qy 63 LARGLSPAFLRFGGKRTDFL-----QFQNLNPAKSRG-----GPGPDYILKNYEDD 109  
Db 255 EEEFL--AHLRGGGRLEQNLAKFSERFDSLRELSAKILGYDSDGETGDFDEEYDDE 310

RESULT 8  
T34697  
probable insertion element ISL647 transposase, truncated - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
C:Accession: T34697  
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, May 1998  
A:Reference number: Z21554  
A:Accession: T34697  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-150 <OLI>  
A:Cross-references: EMBL:AL023702; PIDN:CAAL9229.1; GSPDB:GN00070; SCOEEDB:SC1C3.05  
A:Experimental source: strain A3(2)  
C:Comment: The complete transposase is probably translated by ribosomal frameshifting  
C:Genetics:  
A:Gene: SCOEEDB:SC1C3.05

Query Match 11.9%; Score 74; DB 2; Length 150;  
Best Local Similarity 29.9%; Pred. No. 1.4;  
Matches 29; Conservative 9; Mismatches 33; Indels 26; Gaps 4;  
Qy 5 PLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSKRLVTLA 64  
Db 4 PNPTDR--GKKGSKIHLVDRRGLP-----LSIGISAANLHD-----SQALIPLV 46  
Qy 65 RGLSPAFLRFGGKRTDFLQFQNLNPAKSRGGPGDPY 101  
Db 47 RGIPPTIRSRGPR-----RRPGKLHGDKGYDY 74

RESULT 9  
AF3625  
cell division inhibitor minC [imported] - Brucella melitensis (strain 16M)  
C:Species: Brucella melitensis  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 15-Feb-2002  
C:Accession: AF3625  
R:DeIvecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov  
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Let  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit  
A:Reference number: AD3252; PMID:11756688  
A:Accession: AF3625  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-253 <KUR>  
A:Cross-references: GB:AE008918; PIDN:AAL54169.1; PID:g17985135; GSPDB:GN00191  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BMEI10927  
A:Map position: II  
C:Superfamily: cell division inhibitor minC

Query Match 11.8%; Score 73.5; DB 2; Length 253;  
Best Local Similarity 37.3%; Pred. No. 3;  
Matches 22; Conservative 7; Mismatches 21; Indels 9; Gaps 2;  
Qy 29 PVRTVNENFLSLQLDPSIIHDGWLDFLSSKRLVTLARGLSPAFLRFGGKRTDFLQFQNL 87  
Db 15 PIRLGRSFLAWVLSPELIDGWL-----RLDDLARRSSGFPL----GRPVLDMLNL 64

RESULT 10  
T12095  
beta-fructofuranosidase (EC 3.2.1.26), cell wall - fava bean  
N:Alternate names: cell wall invertase II  
C:Species: Vicia faba (fava bean)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000

Tue Jul 30 10:12:03 2002

us-09-836-461-2\_copy\_42\_161.rpr

C:Accession: T12095  
R:Weber, H.; Borisjuk, L.; Heim, U.; Buchner, P.; Wobus, U.  
Plant Cell 7, 1835-1846, 1995  
A:Title: Seed coat-associated invertases of Fava bean control both unloading and storage  
A:Reference number: Z17416; MUID:96093423  
A:Accession: T12095  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-561 <WEB>  
A:Cross-references: EMBL:Z35163; NID:g861156; PIDN:CAA84527.1; PID:g861157  
A:Experimental source: strain minor; cultivar Frigo; seed coat; clone VfcWINV2  
C:Genetics:  
A:Gene: CWINV2  
C:Function:  
A:Description: hydrolyzes terminal non-reducing beta-D-fructofuranoside residues in beta  
C:Superfamily: beta-fructofuranosidase  
C:Keywords: cell wall; glycoprotein; glycosidase; hydrolase

Query Match 11.8%; Score 73.5; DB 2; Length 581;  
Best Local Similarity 29.9%; Pred. No. 8.8;  
Matches 35; Conservative 19; Mismatches 32; Indels 31; Gaps 10;

QY 25 STNPV-RTVNFNLSIQ--LDPS---LIHDG-WLDFLSSKRLVTLARGLSPAFLRFGK 77  
| : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 168 SPENVIEPTSEKNKINASSFRDPTTGWLGKDGKWRGIVGSKR---STRGIALY-----K 219  
| : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 78 RTDFLOFQNLNPAKSRGGPG---PDYY--LKN-----YEDDIVRS--DVALD 118  
| : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 220 SKDFVKKKSKHPLSAKGTGMWECDFPVLKNGIKGVDTSLNDDYVRHLKVSID 276  
| : ||| : : : : : ||| : : : : : ||| : : : : :  
RESULT 11  
T37000  
probable transposase truncated [imported] - Streptomyces coelicolor (fragment)  
C:Species: Streptomyces coelicolor  
C:Date: 08-Sep-2000 #sequence\_revision 08-Sep-2000 #text\_change 08-Sep-2000  
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, August 1999  
A:Reference number: Z21618  
A:Accession: T37000  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-160 <OLI>  
A:Cross-references: EMBL:AL109949; PIDN:CAB52914.1; GSPDB:GNO0070; SCOEDB:SCJ11.29  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SCJ11.29

Query Match 11.6%; Score 72; DB 2; Length 160;  
Best Local Similarity 29.9%; Pred. No. 2.5;  
Matches 29; Conservative 9; Mismatches 33; Indels 26; Gaps 4;

QY 5 PLPVDRAAGLKEKTLILLDVSTKNPVRTVNFNLSQLQDP--SIHDGWLDFLSSKRLVTLA 64  
| : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 10 PNPVDR--GKKGSKHILITERGLP-----LSIGVSAANTHD-----SQGMEPLV 52  
| : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 65 RGLSPAFLRFGGKRTDFLOFQNLNPAKSRGGPGDY 101  
| : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 53 RGIPPTCSRRGPRR-----RRPAKLHADKGYDY 80  
| : ||| : : : : : ||| : : : : : ||| : : : : :  
RESULT 12  
T09171  
ribosomal protein L4 - spinach  
C:Species: Spinacia oleracea (spinach)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 20-Jun-2000  
C:Accession: T09171  
R:Trifa, Y.; Privat, I.; Gagnon, J.; Baeza, L.; Lerbs-Mache, S.  
J. Biol. Chem. 273, 3980-3985, 1998  
A:Title: The nuclear RPL4 gene encodes a chloroplast protein that co-purifies with the

A:Reference number: Z16600; MUID:98129805  
A:Accession: T09171  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-293 <TRI>  
A:Cross-references: EMBL:X93160  
A:Experimental source: cultivar Melody  
C:Genetics:  
A:Gene: rpl4  
A:Genome: nuclear  
C:Function:  
A:Description: seems to play a role in plastid transcriptional regulation  
C:Superfamily: Escherichia coli ribosomal protein L4  
C:Keywords: chloroplast; protein biosynthesis; ribosome

Query Match 11.6%; Score 72; DB 2; Length 293;  
Best Local Similarity 29.6%; Pred. No. 5.3;  
Matches 29; Conservative 14; Mismatches 31; Indels 24; Gaps 4;

QY 5 PLPVDRAAGLKEKTLILLDVSTKNPVRTVNFNLSQLQDP--SIHDGWLDFLSSKR 59  
| : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 54 PLPILNFSGKE-----VGTEFLNLKTAPPEKARAVVHRLGTLHLQNK 96  
| : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 60 LVTTLARGLSPAFLRFGGKRTDFLOFQNLNPAKSRGGP 97  
| : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 97 RGT-ASTLTRAQVR--GGGRKPYQKKTGRARRSGQSP 132  
| : ||| : : : : : ||| : : : : : ||| : : : : :  
RESULT 13  
T09170  
ribosomal protein L4, chloroplast - spinach  
C:Species: Spinacia oleracea (spinach)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 20-Jun-2000  
C:Accession: T09170  
R:Trifa, Y.; Privat, I.; Gagnon, J.; Baeza, L.; Lerbs-Mache, S.  
J. Biol. Chem. 273, 3980-3985, 1998  
A:Title: The nuclear RPL4 gene encodes a chloroplast protein that co-purifies with the  
A:Reference number: Z16600; MUID:98129805  
A:Accession: T09170  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-293 <TRI>  
A:Cross-references: EMBL:Y14932  
C:Genetics:  
A:Gene: rpl4  
A:Genome: nuclear  
A:Introns: 286/1  
C:Superfamily: Escherichia coli ribosomal protein L4  
C:Keywords: chloroplast; protein biosynthesis; ribosome

Query Match 11.6%; Score 72; DB 2; Length 293;  
Best Local Similarity 29.6%; Pred. No. 5.3;  
Matches 29; Conservative 14; Mismatches 31; Indels 24; Gaps 4;

QY 5 PLPVDRAAGLKEKTLILLDVSTKNPVRTVNFNLSQLQDP--SIHDGWLDFLSSKR 59  
| : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 54 PLPILNFSGKE-----VGTEFLNLKTAPPEKARAVVHRLGTLHLQNK 96  
| : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 60 LVTTLARGLSPAFLRFGGKRTDFLOFQNLNPAKSRGGP 97  
| : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 97 RGT-ASTLTRAQVR--GGGRKPYQKKTGRARRSGQSP 132  
| : ||| : : : : : ||| : : : : : ||| : : : : :  
RESULT 14  
A84206  
hypothetical protein pora [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: A84206  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky  
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja



Jung, K.H.; Alam, M.; Freitas, T.  
PROC. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483  
A:Accession: A84206  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-582 <STO>  
A:Cross-references: GB:AE004437; NID:g10580081; PIDN:AAG19013.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: porA  
C:Superfamily: Halobacterium halobium 2-oxoacid ferredoxin oxidoreductase; 2-oxoacid fer

Query Match 11.6%; Score 72; DB 2; Length 582;  
Best Local Similarity 27.2%; Pred. NO. 13;  
Matches 31; Conservative 17; Mismatches 26; Indels 40; Gaps 5;

QY 8 VDRAAGLKEKTLILLDVSTKNPVRTVNNENFLSLQLDPSIIHGWLDLFLSSKRLVTLARGL 67  
||: : : : ||: : : ||: ||: | | ||: |  
Db 58 VDQVSVVDRDLIALTE-----RTVDENLDELHADSIIYD----- 95

QY 68 SPAFLRFGGKRTDFLQFONLRNPAKSGGPGDYLYLKNYEDD-----IVRSDFVAL 117  
| ||: | | : : ||: | | ||: : | ||: |||  
Db 96 -----GDRTEPADFE---SPAEV---TGLDIPLKDLAEDAGGAIMRNIVAL 135

RESULT 15  
S52500  
SWH1 protein homolog YDL019c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein D2845  
C:Species: Saccharomyces cerevisiae  
C:Date: 08-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 20-Sep-1999  
C:Accession: S52500; S67551  
R:Andre, B.; Vissers, S.; Urrestarazu, L.  
submitted to the EMBL Data Library, February 1995  
A:Description: The sequence of a 42 kb segment located on the left arm of chromosome IV  
A:Reference number: S52492  
A:Accession: S52500  
A:Molecule type: DNA  
A:Residues: 1-1283 <AND>  
A:Cross-references: EMBL:Z48432; NID:g683669; PIDN:CAA88340.1; PID:g683678  
A:Experimental source: strain S288C  
R:Urrestarazu, L.A.; Andre, B.; Vissers, S.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67535  
A:Accession: S67551  
A:Molecule type: DNA  
A:Residues: 1-1283 <URR>  
A:Cross-references: EMBL:Z74067; NID:g1430987; PIDN:CAA98578.1; PID:e253194; PID:g143098  
A:Experimental source: strain S288C  
C:Genetics:  
A:Map position: 4L  
C:Superfamily: unassigned ankryrin repeat proteins; ankryrin repeat homology; EGF homology  
C:Keywords: transmembrane protein  
F:56-72/Domain: transmembrane #status predicted <TMM>

Query Match 11.5%; Score 71.5; DB 2; Length 1283;  
Best Local Similarity 26.0%; Pred. NO. 40;  
Matches 32; Conservative 22; Mismatches 46; Indels 23; Gaps 6;

QY 12 AGLKKEKTLILLDVSTKNPVRTVNNENFLSLQ-LDP-----SIHGWLDLFLSSKRLVTLARG 66  
| : : : || ||| : : ||: ||: | | ||: | : : :  
Db 174 AAMNKRDFSKLDSIWKNP---RNLLDINGIDPETGTTLLYE-----YSQKNDIEMCQW 225

QY 67 L-----SPAFLRFGGKRTDFLQFONLRNPAKSGGPGDYLYLKNY-----EDDIVRSOVA 116  
| : : : | : : | : : ||: | : : |||  
Db 226 LLKHGAEAATVKDGRSPDLDLVKNIKLPKPSNNVTPEIKLNLEKLNREQAIVHEDVA 285

QY 117 LDK 119  
|

Db 286 SSK 288

Search completed: July 30, 2002, 08:16:29  
Job time: 346 sec

us-09-836-461-2\_copy\_42\_161.rpr

Tue Jul 30 10:12:03 2002

\_\_\_\_\_

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:31:30 ; Search time 36.14 Seconds  
(without alignments)  
128.565 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_42\_161  
Perfect score: 621  
Sequence: 1 GORRPLPVDRAAGLKEKTLI.....YLLKNYEDDIVRSDVALDKQ 120

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77.5	12.5	481	1	YFL4_YEAST
2	76	12.2	332	1	W713_CHLMU
3	75	12.1	329	1	Y429_CHLTR
4	72	11.6	293	1	RR4_SPTOL
5	71	11.4	897	1	SVI_FERPE
6	70.5	11.4	992	1	DLP4_RAT
7	70	11.3	1234	1	PIP3_HUMAN
8	68.5	11.0	613	1	CGAA_CLOB1
9	68	11.0	900	1	SYA_MYCGE
10	68	11.0	1067	1	HGBB_HAEIN
11	68	11.0	1084	1	HGP3_HAEIN
12	67	10.8	599	1	CENB_HUMAN
13	67	10.8	742	1	KI21_STRPU
14	66	10.6	367	1	FLC2_PROMI
15	66	10.6	599	1	CENB_MOUSE
16	64.5	10.4	251	1	FLGF_SALTY
17	64.5	10.4	296	1	YF23_HAEIN
18	64.5	10.4	989	1	DLP4_HUMAN
19	64	10.3	274	1	TL30_ARATH
20	64	10.3	345	1	SAE2_YEAST
21	64	10.3	363	1	R23A_HUMAN
22	64	10.3	598	1	HEXA_ALTSO
23	64	10.3	750	1	PSAA_MESVI
24	63.5	10.2	491	1	CAT2_MAIZE
25	63.5	10.2	492	1	ATPB_ANTFO
26	63	10.1	219	1	C270_MOUSE
27	63	10.1	372	1	3BH3_MOUSE
28	63	10.1	671	1	VINE_HUMAN
29	63	10.1	987	1	SOXA_RHIME
30	62.5	10.1	476	1	ATPB_GALSU
31	62.5	10.1	498	1	ATPB_NICPL
32	62.5	10.1	498	1	ATPB_NICRP
33	62.5	10.1	498	1	ATPB_NICSP

34	62.5	10.1	498	1	ATPB_TOBAC	P00826 nicotiana t
35	62.5	10.1	575	1	UL87_EBV	P25215 epstein-bar
36	62.5	10.1	732	1	ACET_HUMAN	P22966 homo sapien
37	62.5	10.1	1054	1	POL_SIVMK	P05897 simian immu
38	62.5	10.1	1118	1	UBP8_HUMAN	P40818 homo sapien
39	62.5	10.1	1306	1	ACE_HUMAN	P12821 homo sapien
40	62	10.0	363	1	R23A_MOUSE	P54726 mus musculu
41	62	10.0	369	1	C5S5_MOUSE	O35926 mus musculu
42	62	10.0	382	1	TP6A_PYRHO	O59209 pyrococcus
43	62	10.0	606	1	CENB_CRIGR	P48988 cricetus
44	61.5	9.9	299	1	LAXC_LACLA	P42096 lactococcus
45	61.5	9.9	299	1	LAXP_LACLA	P23496 lactococcus

ALIGNMENTS

RESULT 1						
YFL4_YEAST						
ID	YFL4_YEAST	STANDARD;	PRT;	481 AA.		
AC	P43616;					
DT	01-NOV-1995 (Rel. 32, Created)					
DT	01-NOV-1995 (Rel. 32, Last sequence update)					
DT	01-OCT-1996 (Rel. 34, Last annotation update)					
DE	Hypothetical 52.9 kDa protein in SAPI55-YMR31 intergenic region.					
GN	YFR044C.					
OS	Saccharomyces cerevisiae (Baker's yeast).					
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;					
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.					
OX	NCBI_TaxID=4932;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=S288C / AB972;					
RX	MEDLINE=95400292; PubMed=7670463;					
RA	Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,					
RA	Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,					
RA	Yamazaki M., Tashiro H., Eki T.;					
RT	"Analysis of the nucleotide sequence of chromosome VI from					
RT	Saccharomyces cerevisiae."					
RL	Nat. Genet. 10:261-268(1995).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=S288C / AB972;					
RX	MEDLINE=96287652; PubMed=8686379;					
RA	Sasanuma M., Tsuchiya Y., Shibata T., Hanaoka F., Murakami Y.,					
RA	Eki T., Naitou M., Hagiwara H., Ozawa M., Sasanuma S.-I.,					
RT	"Analysis of a 36.2 kb DNA sequence including the right telomere of					
RT	chromosome VI from Saccharomyces cerevisiae."					
RL	Yeast 12:149-167(1996).					
CC	-!- SIMILARITY: STRONG, TO THE C-TERMINAL HALF OF YEAST YBR281C.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch).					
CC	-----					
DR	EMBL; D50617; BAA09283.1; -					
DR	EMBL; D44597; BAA08010.1; -					
DR	SGD; S0001940; YFR044C.					
DR	InterPro; IPR002933; Peptidase_M20.					
DR	Pfam; PF01546; Peptidase_M20; 1.					
KW	Hypothetical protein.					
SQ	SEQUENCE 481 AA; 52871 MW; 3E53773A945F5EBC CRC64;					

Query Match 12.5%; Score 77.5; DB 1; Length 481;  
Best Local Similarity 27.8%; Pred No. 0.98;  
Matches 32; Conservative 12; Mismatches 38; Indels 33; Gaps 5;  
QY 17 KTLILLDVSTKNPVRTV-----NENFLSL-----QLDPSIIHDG--WL-- 52

```
DB 334 KTVIPAKVFGFSIRTPDMDSEKLTSLVQKHDAKFKSLNPNKCRTELLHDGAYWSD 393
QY 53 -----DFLSKRLVTLARGLSPAFLRFGGKRTDFLQFQNLNPA-----KSRGGFG 98
DB 394 PNAOFTAAKATKLVYGVDPDFREGGSIPITLTTFQDALNTSVLLLPMDGRGDDG 448

RESULT 2
Y713_CHLMU
ID Y713_CHLMU STANDARD; PRT; 332 AA.
AC Q9PW3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein TC0713.
GN TC0713
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nigg.
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F., Baas S.,
  White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bowan C.,
  Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
  RA Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Sdzberg S.L.,
  RA Eisen J., Fraser C.M.;
  RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
  RT pneumoniae AR39 "; 28:1397-1406(2000).
RL Nucleic Acids Res.
CC -1- SIMILARITY: BELONGS TO THE UPF0158 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001316; AAC68026.1;
CC DR Hypothetical protein; Complete proteome.
CC KW DOMAIN 301 316 ASP/GLU-RICH (ACIDIC).
CC FT SEQUENCE 329 AA; 39263 MW; B1706123BE7F7698 CRC64;
CC SQ

Query Match 12.2%; Score 76; DB 1; Length 332;
Best Local Similarity 28.8%; Pred. No. 0.89; Indels 20; Gaps 6;
Matches 34; Conservative 19; Mismatches 45;

QY 3 RRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNFNLSLQDPSIHDGWLDFLSSKRLVT 62
DB 202 KRGPRPKQSAKVEATTSSDIYTKVP--SAARREFLF---PEITSASSLTF--SEKFT 254
QY 63 LARGLSPAFLRFGGKRTDFL-----QFONLRNPAKSRG----GPGPDYLNKYEDD 109
DB 255 EEEFL--AHLRGGGRLEQDLNLAKEFSERFDSLRELSAKLGYDSDGDTGDFDEEYDD 310

RESULT 4
RK4_SPIOL
ID RK4_SPIOL STANDARD; PRT; 293 AA.
AC Q49937; Q49938;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L4, chloroplast precursor (R-protein L4).
GN RPL4.
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Chenopodiaceae; Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 51-63.
RC STRAIN=CV. MELODY; TISSUE=Leaf;
RX MEDLINE=98129805; PubMed=9461586;
RA Trifa V., Privat I., Gagnon J., Baeza L., Lerbs-Mache S.;
RT "The nuclear RPL4 gene encodes a chloroplast protein that co-purifies
  RT with the 77-like transcription complex as well as plastid
  RT ribosomes.";
  RL J. Biol. Chem. 273:3980-3985(1998).
RN [2]
RP SEQUENCE OF 51-68; 82-86; 98-105; 131-137; 152-158; 211-224 AND
  RP 228-235.
RC STRAIN=CV. ALVARO; TISSUE=Leaf;
RX MEDLINE=20435798; PubMed=10874046;
RA Yamaguchi K., Subramanian A.R.;
RT "The plastid ribosomal proteins. Identification of all the proteins in
  RT the 50 S subunit of an organelle ribosome (chloroplast).";
  RL J. Biol. Chem. 275:28466-28482(2000).
```

```
DB 334 KTVIPAKVFGFSIRTPDMDSEKLTSLVQKHDAKFKSLNPNKCRTELLHDGAYWSD 393
QY 53 -----DFLSKRLVTLARGLSPAFLRFGGKRTDFLQFQNLNPA-----KSRGGFG 98
DB 394 PNAOFTAAKATKLVYGVDPDFREGGSIPITLTTFQDALNTSVLLLPMDGRGDDG 448

RESULT 2
Y713_CHLMU
ID Y713_CHLMU STANDARD; PRT; 332 AA.
AC Q9PW3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein TC0713.
GN TC0713
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nigg.
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F., Baas S.,
  White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bowan C.,
  Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
  RA Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Sdzberg S.L.,
  RA Eisen J., Fraser C.M.;
  RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
  RT pneumoniae AR39 "; 28:1397-1406(2000).
RL Nucleic Acids Res.
CC -1- SIMILARITY: BELONGS TO THE UPF0158 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE002339; AAF39526.1;
CC DR TIGR; TC0713;
CC KW Hypothetical protein; Complete proteome.
CC FT DOMAIN 301 316 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 308 316 POLY-ASP.
CC SQ SEQUENCE 332 AA; 39202 MW; 27527DC5E6ABDEA4 CRC64;

Query Match 12.2%; Score 76; DB 1; Length 332;
Best Local Similarity 28.8%; Pred. No. 0.89; Indels 20; Gaps 6;
Matches 34; Conservative 19; Mismatches 45;

QY 3 RRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNFNLSLQDPSIHDGWLDFLSSKRLVT 62
DB 202 KRGPRPKQSAKVEATTSSDIYTKVP--SAARREFLF---PEITSASSLTF--SEKFT 254
QY 63 LARGLSPAFLRFGGKRTDFL-----QFONLRNPAKSRG----GPGPDYLNKYEDD 109
DB 255 EEEFL--AHLRGGGRLEQDLNLAKEFSERFDSLRELSAKLGYDSDGDTGDFDEEYDD 310

RESULT 3
Y429_CHLTR
ID Y429_CHLTR STANDARD; PRT; 329 AA.
AC O84436;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein CT429.
GN CT429.
OS Chlamydia trachomatis.
```





Query Match 11.3%; Score 70; DB 1; Length 1234;  
Best Local Similarity 29.6%; Pred. No. 20;  
Matches 32; Conservative 18; Mismatches 44; Indels 14; Gaps 4;

[illegible]

RESULT	8
CLOBI	
ID	CGAA_CLOBI
STANDARD;	PRT; 613 AA.
AC	Q45882;
DT	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Pesticidal crystal-like protein cryl6Aa (Insecticidal toxin
DE	CryXIV(a)) (Cbm71 mosquitoctocidal toxin).
GN	CryI6AA OR CryXIV(A) OR CBM71.
OS	Clostridium bifermentans.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC	Clostridium.
OX	NCBI_TaxID=1490;

11.1	SEQUENCE FROM N.A., AND SEQUENCE OF 1-15 AND 374-388.
RP	STRAIN=CH18 / SUBSP. MALAYSIA;
RC	MEDLINE=962336023; PubMed=8655486;
RA	Barloy F., Delecluse A., Nicolas L., Lecadet M.-M.;
RT	"Cloning and expression of the first anaerobic toxin gene from
RT	<i>Clostridium bifermentans</i> subsp. <i>malaysia</i> , encoding a new mosquitocidal
RT	protein with homologies to <i>Bacillus thuringiensis</i> delta-endotoxins.";
J.	<i>Bacteriol.</i> 178:3099-3105(1996).
RL	-!- FUNCTION: TOXIN ACTIVE ON MOSQUITO LARVAE OF THE SPECIES AEDES
CC	AEGYPTI, CULEX PIPIENS, AND ANOPHELES STEPHENSI.
CC	-!- SUBCELLULAR LOCATION: Secreted.
CC	-!- DEVELOPMENTAL STAGE: PRODUCED DURING SPORULATION. THE EXPRESSION
CC	DECREASES CONSIDERABLY WITH CELL LYSIS.
CC	-!- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR	EMBL; X941146; CAA63860.1; -
DR	InterPro; IPR001178; Endotoxin.
DR	Fram; PF00355; endotoxin; 1.
DR	Toxin; Sporulation.
KW	DOMAIN 155 160 POLY-LEU.
FT	CONFILCT 9 9 H -> Q (IN REF. 1; AA SEQUENCE).
SQ	SEQUENCE 613 AA; 71172 MW; 410EFDD9C601CEE5 CRC64;
ST	

```

Query Match          11.0%; Score 68.5; DB 1; Length 613;
Best Local Similarity 25.6%; Pred. No. 12;
Matches 23; Conservative 15; Mismatches 39; Indels 13; Gaps 2;

QY 19 LILLVSTKNPVRTVNENFSLQDPSIIHGWGDFLSSRLVLTARGLSPAFLRFGGKR 78
||| : : : : : ||| : : : : :
Db 533 LIRIAGSKYITNIQOTFNTENNPSLIYDDFKYFNFETLSITSSIDLILEFYYSY 592

QY 79 TD--FLQFQNLRNPAKSRGGPGDYLYKNY 106
||| : : : : : ||| : : : : :
Db 593 TDGNFEDFPKLSIP-----YRNY 611

```

RESULT 9  
SYA MYCGE

ID	SYA_MYCGE	STANDARD	PRT	900 AA
AC	P47534;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Alanine-tRNA synthetase (EC 6.1.1.7) (Alanine-tRNA ligase) (Alars).			
GN	ALAS OR MG292.			
OS	Mycoplasma genitalium.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;			
OC	Mycoplasmataceae; Mycoplasma.			
OX	NCBI_TaxID=2097;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=ATCC 33530 / G-37;			
RX	MEDLINE=9602346; PubMed=7569993;			
RA	Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,			
RA	Fleischmann R.D., Bult C.J., Krollavage A.R., Sutton G., Kelley J.M.,			
RA	Frishman J.L., Weidman J.F., Saudek D.M., Phillips C.A., Merrick J.L.,			
RA	Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.L.,			
RA	Tomb J.F., Dougherty B.A., Bost K.F., Hu P.-C., Lucier T.S.,			
RA	Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;			
RL	"The minimal gene complement of Mycoplasma genitalium.";			
RL	Science 270:397-403(1995).			

CC -I- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) -> AMP +  
CC diphosphate + L-alanyl-tRNA(Ala).  
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -I- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

```

CC -----
DR ENBL; U39709; AAC71513.1; -
DR TIGR; MG292; -
DR InterPro; IPR002106; AA_trna_ligase_II.
DR InterPro; IPR002318; trna-synt_2c.
DR Pfam; PFO1411; trna-synt_2c; 1.
DR PRINTS; PR00980; TRNASYNTHALA.
DR PROSITE; PS001179; AA_TRNA_LIGASE_II_1; FALSE_NEG.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; 1.
DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 900 AA; 104301 MW; AA54520BFB3949A2 CRC64;

```

Query Match	11.0%;	Score 68;	DB 1;	Length 900;
Best Local Similarity	21.2%;	Pred. No. 22;		

```

Qy 11 AAGLKEKTLILLDVSTKPNVTVNENFLSLQDPSI---IHDGWLDFLSKKRLVTLAG 66
Db 793 ALSIKQLSLVDENKSKYVATFND---VEPKLLQLTHDVFNQNTKFNLIINQFN 846

```

**Qy** 67 LSPAELRFGSGKRTDFLQFQNLRNPAKSRGGPGDPYVLYKNYEDDI 110  
:  
:  
:  
**Dd** 847 ESNSEIVIGKNTTTTII--EKLRSNPLKGGGENDKLFGRGSQDNV 888  
:  
:  
:

RESULT	10
HGBB_HAEIN	
ID	HGBB_HAEIN
AC	Q9KIV1;
DT	16-OCT-2001
DT	(Rel. 40, Created)
DT	16-OCT-2001
DT	(Rel. 40, Last sequence update)
DT	16-OCT-2001
DT	(Rel. 40, Last annotation update)

1 : : : : :  
711 YKQNFNDL 719

Db

RESULT 11  
HGP3\_HAEIN STANDARD; PRT; 1084 AA.  
AC P44836;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Probable hemoglobin and hemoglobin-haptoglobin binding protein 3  
DE precursor.  
GN HI0712.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Karlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA Mckenney K., Sutton G., Liu L.-I., Glodek A., Kelley J.M.,  
RA Scott J.D., Shirley R., Liu L.-I., Fields C.A., Gocayne J.D.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT influenzae Rd.";  
RL Science 269:496-512(1995).  
[2]  
RN IDENTIFICATION BY MASS SPECTROMETRY.  
RP MEDLINE=20137488; PubMed=10675023;  
RX Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,  
RA Gray C., Fountoulakis M.;  
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";  
RA Electrophoresis 21:411-429(2000).  
CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE  
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED  
CC FOR HEME UPTAKE (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).  
CC -!- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE  
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA  
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.  
CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING  
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE  
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A WAY TO  
CC AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR FAMILY;  
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U32754; AAC22369.1; -  
CC TIGR; HI0712; -  
DR InterPro; IPR000531; TonB\_boxC.  
DR Pfam; PF00593; TonB\_boxC; 1.  
DR PROSITE; PS00430; TONB\_DEPENDENT\_REC.1; FALSE\_NEG.  
DR PROSITE; PS01156; TONB\_DEPENDENT\_REC.2; 1.  
KW Outer membrane; Transport; TonB box; Multigene family; Signal;  
KW Receptor; Repeat; Complete proteome.  
FT SIGNAL 1 24 POTENTIAL.  
FT CHAIN 25 1084 PROBABLE HEMOGLOBIN AND HEMOGLOBIN-



FT DOMAIN 26 73 HAPTOGLOBIN BINDING PROTEIN 3.  
 FT REPEAT 26 29 12 X 4 AA TANDEM REPEATS OF Q-P-T-N.  
 FT REPEAT 30 33 1.  
 FT REPEAT 34 37 2.  
 FT REPEAT 38 41 3.  
 FT REPEAT 42 45 4.  
 FT REPEAT 46 49 5.  
 FT REPEAT 50 53 6.  
 FT REPEAT 54 57 7.  
 FT REPEAT 58 61 8.  
 FT REPEAT 62 65 9.  
 FT REPEAT 66 69 10.  
 FT REPEAT 70 73 11.  
 FT SITE 83 90 TONB BOX.  
 FT SITE 1067 1084 TONB C-TERMINAL BOX.  
 FT SITE 1067 1084 TONB C-TERMINAL BOX.  
 FT SEQUENCE 1084 AA; 123955 MW; 794DF91E0F53CFD9 CRC64;

Query Match 11.0%; Score 68; DB 1; Length 1084;  
 Best Local Similarity 26.9%; Pred. No. 28;  
 Matches 35; Conservative 16; Mismatches 47; Indels 32; Gaps 8;

QY 1 GDRPLP-VDRAGL-----KEKTLILDVSTKNPVRTVNN-----FLSLQDPSIIHDG 50  
 Db 619 GKPTAPDPYSRCSLMNTDQKETYLIPTVTKNNVLYFGDNVQLTSWLGDLNRYDHVK 678  
 QY 51 WLDFLSKRLVTLARGL-SPAFLRFGKGTDFLQFNLRNPAKSGRGGPDY-----101  
 Db 679 YLPSYDEK--IPVNGLLTGLFKKFGPK--DIVYGSKYSKPA-----DYDCTYNSD 726  
 QY 102 -YLNKYEDDI 110  
 Db 727 CYKKNFKDNL 736

RESULT 12  
 CENB\_HUMAN STANDARD; PRT; 599 AA.  
 AC P07199;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Major centromere autoantigen B (Centromere protein B) (CENP-B).  
 GN CENPB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91372020; PubMed=1893793;  
 RA Sullivan K.F., Glass C.A.;  
 RT "CENP-B is a highly conserved mammalian centromere protein with  
 RT homology to the helix-loop-helix family of proteins.";  
 RL Chromosoma 100:360-370(1991).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguely C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dharm P., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Leinvaslao M.H., Leversha M., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,

RA Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 [3]  
 RP SEQUENCE OF 6-599 FROM N.A.  
 RX MEDLINE=87166180; PubMed=2435739;  
 RA Earnshaw W.C., Sullivan K.F., Machlin P.S., Cooke C.A.,  
 RA Kaiser D.A., Pollard T.D., Rothfield N.F., Cleveland D.W.;  
 RT "Molecular cloning of cDNA for CENP-B, the major human centromere  
 RT autoantigen.";  
 RL J. Cell Biol. 104:817-829(1987).  
 [4]  
 RP SUBUNITS, AND DOMAINS.  
 RX MEDLINE=93107144; PubMed=1469042;  
 RA Yoda K., Kitagawa K., Masumoto H., Muro Y., Okazaki T.;  
 RT "A human centromere protein, CENP-B, has a DNA binding domain  
 RT containing four potential alpha helices at the NH2 terminus, which is  
 RT separable from dimerizing activity.";  
 RL J. Cell Biol. 119:1413-1427(1992).  
 [5]  
 RP STRUCTURE BY NMR OF 1-56.  
 RX MEDLINE=98119825; PubMed=9451007;  
 RA Iwahara J., Kigawa T., Kitagawa K., Masumoto H., Okazaki T.,  
 RA Yokoyama S.;  
 RT "A helix-turn-helix structure unit in human centromere protein B  
 RT (CENP-B).";  
 RL EMBO J. 17:827-837(1998).  
 [6]  
 RP FUNCTION: INTERACTS WITH CENTROMERIC HETEROCHROMATIN IN  
 RP CHROMOSOMES AND BINDS TO A SPECIFIC SUBSET OF ALPHOID SATELLITE  
 RP DNA, CALLED THE CENP-B BOX. MAY ORGANIZE ARRAYS OF CENTROMERE  
 RP SATELLITE DNA INTO A HIGHER ORDER STRUCTURE WHICH THEN DIRECTS  
 RP CENTROMERE FORMATION AND KINETOCORE ASSEMBLY IN MAMMALIAN  
 RP CHROMOSOMES.  
 [7]  
 RP SUBUNIT: HOMODIMER.  
 [8]  
 RP SUBCELLULAR LOCATION: Nuclear.  
 [9]  
 RP SIMILARITY: STRONG, WITH OTHER MAMMALIAN CENP-B.  
 [10]  
 RP This SWISS-PROT entry is copyright. It is produced through a collaboration  
 RP between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 RP the European Bioinformatics Institute. There are no restrictions on its  
 RP use by non-profit institutions as long as its content is in no way  
 RP modified and this statement is not removed. Usage by and for commercial  
 RP entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 RP or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 [11]  
 RP EMBL; X55039; CAA38879.1; -;  
 DR EMBL; AL109804; CAC17547.1; -;  
 DR EMBL; X05299; CAA28918.1; -;  
 DR PIR; A27272; A27272;  
 DR PIR; S18735; S18735;  
 DR PIR; I8W6; 07-OCT-98.  
 DR PDB; 1L7140; -;  
 DR MM; 117140; -;  
 KW Chromosomal protein; Nuclear protein; DNA-binding; Centromere;  
 KW 3D-structure.  
 FT DNA\_BIND 1 125  
 FT DOMAIN 404 465 GLU-RICH (ACIDIC)  
 FT DOMAIN 508 538 ASP/GLU-RICH (ACIDIC).  
 FT CONFLICT 583 583 R -> M (IN REF. 3).  
 FT CONFLICT 592 593 VR -> LL (IN REF. 3).  
 SQ SEQUENCE 599 AA; 65171 MW; 9B4B7DB957A914AA CRC64;

Query Match 10.8%; Score 67; DB 1; Length 599;  
 Best Local Similarity 28.7%; Pred. No. 17;

Matches 25; Conservative 10; Mismatches 44; Indels 8; Gaps 2;	
Qy 10	RAAGLKEKTLILLDVSTKNPVRTVNFNLSLQLDPSIIHDGWLDFLSKRLVTLARGLSP 69
Db 89	RAAGLPVKGIIL---KEKALRIAE---LGMDDFTASNGWLDLFRFRRRHGVVSCSGVAR 140
Qy 70	AFLRFGGKRTDFLOFQNLNRPNAKSRGG 96
Db 141	ARARNAAPRTPAAPASPAVPEGGG 167
RESULT 13	
KI2L_STRPU	
ID KI2L_STRPU	STANDARD; PRT; 742 AA.
AC P46871;	
DT 01-NOV-1995 (Rel. 32, Created)	
DT 01-NOV-1995 (Rel. 32, Last sequence update)	
DT 15-DEC-1998 (Rel. 37, Last annotation update)	
DE Kinesin-II 95 kDa subunit (KRP-85/95 95 kDa subunit).	
GN KRP95.	
OS Strongylocentrotus purpuratus (Purple sea urchin).	
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;	
OC Echinolidea; Euechinolidea; Echinacea; Echinolida; Strongylocentrotidae;	
OC Strongylocentrotus.	
OX NCBI_TaxID=7668;	
RN [1]	
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.	
RC TISSUE=Egg;	
RA MEDLINE=94050179; PubMed=8232586;	
RA Cole D.G., Chinn S.W., Wedaman K.P., Hall K., Vuong T.,	
RA Scholey J.M.;	
RT "Novel heterotrimeric kinesin-related protein purified from sea	
RT urchin eggs.";	
RL Nature 366:268-270(1993).	
CC -1- SUBUNIT: TRIMER OF A 115 kDa SUBUNIT AND TWO KINESIN-LIKE SUBUNITS	
CC OF 95 AND 85 kDa.	
CC -1- PTM: THE N-TERMINUS IS BLOCKED.	
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN	
CC II SUBFAMILY.	
CC	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	
CC entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	
CC EMBL; U00996; AAA87393.1; -	
DR HSSP; P17119; 3KAR.	
DR InterPro: IPR001752; kinesin.	
DR Pfam; PF00225; kinesin; 1.	
DR PRINTS; PR00380; KINESINHEAVY.	
DR SMART; SM00129; KISC; 1.	
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.	
DR PROSITE; PS00057; KINESIN_MOTOR_DOMAIN2; 1.	
KW Motor protein; Microtubules; ATP-binding; Coiled coil.	
FT DOMAIN 1 337 KINESIN-MOTOR (BY SIMILARITY).	
FT DOMAIN 338 613 COILED COIL (BY SIMILARITY).	
FT DOMAIN 614 742 GLOBULAR (BY SIMILARITY).	
FT NP_BIND 95 102 ATP (POTENTIAL).	
FT SEQUENCE 742 AA; 84202 MW; 47C40A367BAA77B5 CRC64;	
Query Match 10.8%; Score 67; DB 1; Length 742;	
Best Local Similarity 25.8%; Pred. No. 23;	
Matches 34; Conservative 15; Mismatches 53; Indels 30; Gaps 6;	
Qy 1	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNFNLSLQLDPSIIH-----DGWL---- 52
Db 258	GDR-----LKEATKINLSALG-----NVISALVDGKSHPIYRDSKLTRELL 300
Qy 53	DFLSSKRLVTLARGLSPAFLRFGGKRTDFL---QFQNLNRPNAKSRGGPGPDYLLKNYED 108
Matches 25; Conservative 10; Mismatches 44; Indels 8; Gaps 2;	
Qy 109	DIVRSVDALDKQ 120
Db 360	EISRLKQALDDK 371
RESULT 14	
FLC2_PROMI	
ID FLC2_PROMI	STANDARD; PRT; 367 AA.
AC P42273;	
DT 01-NOV-1995 (Rel. 32, Created)	
DT 01-NOV-1995 (Rel. 32, Last sequence update)	
DT 16-OCT-2001 (Rel. 40, Last annotation update)	
DE Flagellin 2.	
GN FLIC2.	
OS Proteus mirabilis.	
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;	
OC Proteus.	
OX NCBI_TaxID=584;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=BB2000;	
RX MEDLINE=95011656; PubMed=7926835;	
RA Belas R., Flaherty D.;	
RT "Sequence and genetic analysis of multiple flagellin-encoding genes	
RT from Proteus mirabilis.";	
RL Gene 148:33-41(1994).	
CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO	
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.	
CC -1- INDUCTION: ALTHOUGH SWIMMER CELLS HAVE ONLY A FEW FLAGELLA, THE	
CC ELONGATED SWARMER CELLS ARE PROFUSELY COVERED BY THOUSANDS OF NEW	
CC FLAGELLA SYNTHESIZED SPECIFICALLY IN RESPONSE TO GROWTH ON	
CC SURFACES OR IN HIGHLY VISCOUS LIQUIDS.	
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.	
CC	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	
CC entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	
CC EMBL; AF221596; AAA62397.1; -	
DR InterPro: IPR001029; Flagellin_C.	
DR InterPro: IPR001492; Flagellin_N.	
DR Pfam; PF00700; Flagellin_C; 1.	
DR Pfam; PF00669; Flagellin_N; 1.	
DR PRINTS; PR00207; FLAGELLIN.	
DR ProDom; PD000316; Flagellin_C; 1.	
KW Flagella; Multigene family.	
Qy 43	DPSIIHDGWLDFLSKRLVTLARGLSPAFLRFGGKRTDFLOFQNLNRPNAKSRGGPGPDY 102
Db 224	KVLDNGSDDIIVSKSDFTLKSGTTGTEVETGSKT-----TKFTADAGKDVK 272
Query Match 10.6%; Score 66; DB 1; Length 367;	
Best Local Similarity 26.9%; Pred. No. 12;	
Matches 21; Conservative 13; Mismatches 32; Indels 12; Gaps 2;	
Qy 103	LKNYEDDIVRS-DVALDK 119
Db 273	VLNVKDALATLDNAISK 290
RESULT 15	
CENB_MOUSE	
ID CENB_MOUSE	STANDARD; PRT; 599 AA.
AC P27790;	



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:30:45 ; Search time 123.94 Seconds  
(without alignments)  
167.495 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_42\_161  
Perfect score: 621  
Sequence: 1 GDRRLPVDRAAGLKEKTLI.....YLLKNVEDDIVRSDVALDKQ 120

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	621	100.0	534	4 Q9HB38	Q9hb38 homo sapien
2	621	100.0	592	4 Q9HB37	Q9hb37 homo sapien
3	566	91.1	480	4 Q9HB39	Q9hb39 homo sapien
4	163	26.2	536	11 Q9QZF8	Q9qzf8 rattus norv
5	158.5	25.5	545	6 Q9MYX0	Q9myx0 bos taurus
6	157.5	25.4	545	4 Q9UL39	Q9ul39 homo sapien
7	156.5	25.2	543	4 Q9Y251	Q9y251 homo sapien
8	136.5	22.0	523	13 Q90YK5	Q90yk5 gallus gall
9	85.5	13.8	1391	5 O17772	O17772 caenorhabdi
10	84.5	13.6	304	10 Q9C5M5	Q9c5m5 arabidopsis
11	84.5	13.6	334	10 Q9LFY5	Q9lfy5 arabidopsis
12	82.5	13.3	756	16 Q988J0	Q988j0 rhizobium l
13	77	12.4	385	10 Q93X83	Q93x83 vitis labru
14	75	12.1	241	5 Q9W3U3	Q9w3u3 drosophila
15	74	11.9	150	2 Q98847	Q98847 streptomyce
16	73.5	11.8	581	10 Q43856	Q43856 vicia faba

17	73.5	11.8	705	4 Q9H9S9	Q9h9s9 homo sapien
18	72	11.6	160	2 Q9RI77	Q9ri77 streptomyce
19	72	11.6	582	17 Q9HR25	Q9hr25 halobacteri
20	71.5	11.5	527	10 Q9LRC8	Q9lrc8 scutellaria
21	71.5	11.5	1283	3 Q12451	Q12451 saccharomyc
22	71	11.4	881	5 Q9GRI1	Q9gri1 paramecium
23	70.5	11.4	433	2 Q937S1	Q937s1 bruceella me
24	70	11.3	1983	4 Q9Y4D6	Q9y4d6 homo sapien
25	69.5	11.2	378	2 Q939J7	Q939j7 campylobact
26	69.5	11.2	1167	16 Q9A205	Q9a205 streptococc
27	69	11.1	116	4 Q9UH21	Q9uh21 homo sapien
28	69	11.1	116	4 Q9Y3I9	Q9y3i9 homo sapien
29	69	11.1	203	4 Q9NQF3	Q9nqf3 homo sapien
30	69	11.1	314	4 Q9H4I8	Q9h4i8 homo sapien
31	69	11.1	329	10 Q9FGI3	Q9fgi3 arabidopsis
32	69	11.1	363	12 Q73494	Q73494 okra yellow
33	69	11.1	409	15 Q90N44	Q90n44 human immun
34	69	11.1	624	10 Q9FGH9	Q9fgh9 arabidopsis
35	69	11.1	838	10 Q9FG37	Q9fg37 arabidopsis
36	69	11.1	993	2 Q48915	Q48915 microcystis
37	68.5	11.0	378	16 Q9PMY3	Q9pmy3 campylobact
38	68.5	11.0	463	8 Q9MRQ8	Q9mrq8 ceratophyll
39	68.5	11.0	694	16 Q98PE1	Q98pel mycoplasma
40	68	11.0	321	15 Q90KT0	Q90kt0 human immun
41	68	11.0	331	5 P91889	P91889 sepia offic
42	68	11.0	622	11 Q923C5	Q923c5 mus musculu
43	68	11.0	2126	2 Q9S1A8	Q9s1a8 microcystis
44	68	11.0	2126	2 Q93LR2	Q93lr2 microcystis
45	68	11.0	2126	2 Q93LRI	Q93lrl microcystis

ALIGNMENTS

RESULT 1

Q9HB38	Q9HB38	PRELIMINARY:	PRT:	534 AA.
ID	Q9HB38			
AC	Q9HB38			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	HEPARANASE-LIKE PROTEIN HPA2B.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=20483645; PubMed=11027606;			
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,			
RA	Hircock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;			
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian			
RT	Heparanase Family Member.;"			
RL	Biochem. Biophys. Res. Commun. 276:1170-1177(2000).			
DR	EMBL; AF282886; AAG23422.1; -			
SQ	SEQUENCE 534 AA; 60063 MW; C3DE5E900CB338C4 CRC64;			

Query Match 100.0%; Score 621; DB 4; Length 534;  
Best Local Similarity 100.0%; Pred. No. 1.4e-59;  
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNEFNFLSLQLDPSIHDGWLDFLSKRL	60
DB	42	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNEFNFLSLQLDPSIHDGWLDFLSKRL	101
QY	61	VTLARGSLPAFLRFQCKRTDFLQFQNLNPAKSRGGPGPDYLLKNYEDDIVRSDVALDKQ	120
DB	102	VTLARGSLPAFLRFQCKRTDFLQFQNLNPAKSRGGPGPDYLLKNYEDDIVRSDVALDKQ	161

RESULT 2

Q9HB37	Q9HB37
--------	--------

Q9QZF8		PRELIMINARY;	PRT;	536 AA.
AC	Q9QZF8;			
DT	01-MAY-2000 (TReMBLrel. 13, Created)			
DT	01-MAY-2000 (TReMBLrel. 13, Last sequence update)			
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)			
DE	HEPARANASE.			
GN	HSP.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Podyma K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;			
RT	"Heparanase from parathyroid cell line.;"			
RL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF184967; AAF04563.1; -.			
SQ	SEQUENCE 536 AA; 60569 MW; 6208B1FD9EE28421 CRC64;			
Query Match 26.2%; Score 163; DB 11; Length 536;				
Best Local Similarity 39.7%; Pred. No. 2.2e-09;				
Matches 46; Conservative 17; Mismatches 47; Indels 6; Gaps 3;				
QY	1 GDRRLPVDRAAG-LKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHD-GWLDLSSK 58			
Db				
QY	12 GLRLTQGTAGTAPTQDVVDLEFYTKRLFQSVSPSLITIDASLATDPRFLTLGSP 71			
Db				
QY	59 RLVTIARGLSPAFLRFGGKRTDFLOFQNLNPNKSRGGPGDPYLLKNEYDDIVRSD 114			
Db				
QY	72 RLRLARGLSPAFLRFGGKRTDFLIF- - -DPNKEPTSEERSYQSDNNICGSE 123			
Db				
RESULT 5				
Q9MY10	PRELIMINARY;	PRT;	545 AA.	
ID	Q9MY10			
AC	Q9MY10;			
DT	01-OCT-2000 (TReMBLrel. 15, Created)			
DT	01-JUN-2001 (TReMBLrel. 17, Last sequence update)			
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)			
DE	HEPARANASE.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=PLACENTA;			
RA	Kizaki K., Nakano H., Takahashi T., Imal K., Hashizume K.;			
RT	"Expression of Heparanase mRNA in Bovine Placenta During Gestation.;"			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF281160; AAF87301.2; -.			
SQ	SEQUENCE 545 AA; 61076 MW; FAC4BDFD855B933 CRC64;			
Query Match 25.5%; Score 158.5; DB 6; Length 545;				
Best Local Similarity 38.0%; Pred. No. 7e-09;				
Matches 41; Conservative 15; Mismatches 39; Indels 13; Gaps 3;				
QY	7 PVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHD-GWLDLSSKRLVTLAR 65			
Db				
QY	36 PADDAE- - - - -LEFFTERPLHLVSPAFLSFTIDANLATDPRFTFLGSSKRLTLAR 87			
Db				
QY	66 GLSPAFLRFGGKRTDFLOFQNLNPNKSRGGPGDPYLLKNEYDDIVRS 113			
Db				
QY	88 GLAPALRFGNGKDFLIDFPKKEPAEE- - - - -RSYWLQSQSDNODICKS 131			
Db				
RESULT 6				
Q9UL39	PRELIMINARY;	PRT;	545 AA.	
ID	Q9UL39			

Q9HB37		PRELIMINARY;	PRT;	592 AA.
AC	Q9HB37;			
DT	01-MAR-2001 (TReMBLrel. 16, Created)			
DT	01-MAR-2001 (TReMBLrel. 16, Last sequence update)			
DT	01-MAR-2001 (TReMBLrel. 16, Last annotation update)			
DE	HEPARANASE-LIKE PROTEIN HPA2C.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20483645; PubMed=11027606;			
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,			
RA	Hirocock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;			
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian			
RT	Heparanase Family Member.;"			
RL	Biochem. Biophys. Res. Commun. 276:1170-1177(2000).			
DR	EMBL; AF282887; AAG23423.1; -.			
SQ	SEQUENCE 592 AA; 66580 MW; 95C384AD9A74258E CRC64;			
Query Match 100.0%; Score 621; DB 4; Length 592;				
Best Local Similarity 100.0%; Pred. No. 1.6e-59;				
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSKRL 60			
Db				
QY	42 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSKRL 101			
Db				
QY	61 VTLARGLSPAFLRFGGKRTDFLOFQNLNPNKSRGGPGDPYLLKNEYDDIVRSDVALDKQ 120			
Db				
QY	102 VTLARGLSPAFLRFGGKRTDFLOFQNLNPNKSRGGPGDPYLLKNEYDDIVRSDVALDKQ 161			
Db				
RESULT 3				
Q9HB39	PRELIMINARY;	PRT;	480 AA.	
ID	Q9HB39			
AC	Q9HB39;			
DT	01-MAR-2001 (TReMBLrel. 16, Created)			
DT	01-MAR-2001 (TReMBLrel. 16, Last sequence update)			
DT	01-MAR-2001 (TReMBLrel. 16, Last annotation update)			
DE	HEPARANASE-LIKE PROTEIN HPA2A.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20483645; PubMed=11027606;			
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,			
RA	Hirocock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;			
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian			
RT	Heparanase Family Member.;"			
RL	Biochem. Biophys. Res. Commun. 276:1170-1177(2000).			
DR	EMBL; AF282885; AAG23421.1; -.			
SQ	SEQUENCE 480 AA; 53900 MW; F75F89F67AC1FF83 CRC64;			
Query Match 91.1%; Score 566; DB 4; Length 480;				
Best Local Similarity 99.1%; Pred. No. 1.3e-53;				
Matches 108; Conservative 1; Mismatches 0; Indels 0; Gaps 0;				
QY	1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSKRL 60			
Db				
QY	42 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHDGWLDFLSSKRL 101			
Db				
QY	61 VTLARGLSPAFLRFGGKRTDFLOFQNLNPNKSRGGPGDPYLLKNEYDD 109			
Db				
QY	102 VTLARGLSPAFLRFGGKRTDFLOFQNLNPNKSRGGPGDPYLLKNEYDE 150			
Db				
RESULT 4				
Q9UL39	PRELIMINARY;	PRT;	545 AA.	
ID	Q9UL39			



```
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
ON NCBI_TaxID=6239;
RX [1]
RP SEQUENCE FROM N.A.
RA Lennard N.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
   investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z92781; CAB07179.2; -.
DR InterPro: IPR0011715; Calponin_hom.
DR InterPro: IPR003247; CH_type.
DR InterPro: IPR000048; IQ.
DR InterPro: IPR001936; RasGAP.
DR InterPro: IPR000593; RasGAP_C.
DR Pfam: PF00612; IQ; 2.
DR ProDom: PD001527; CH_type; 1.
DR ProDom: PD008735; RasGAP_C; 1.
DR SMART: SM00033; CH; 1.
DR SMART: SM00015; IQ; 1.
DR PROSITE: PS50021; CH; 1.
DR PROSITE: PS50018; RAS_GTPASE_ACTIV_2; 1.
DR PROSITE: PS50018; RAS_GTPASE_ACTIV_2; 1.
SQ SEQUENCE 1391 AA; 159210 MW; 794DF9EBA2E65AC CRC64;

Query Match 13.8%; Score 85.5; DB 5; Length 1391;
Best Local Similarity 23.8%; Pred. No. 2.2;
Matches 35; Conservative 15; Mismatches 52; Indels 45; Gaps 4;

QY 8 VDAAGLKERTLLLDVSTKNP--VRTVNFNLSLQD----- 43
Db 485 VEAFLAEQVVRINKAGSEPKVRELEKLSFKLDGYREELHWHYAKRISEELDASKH 544

QY 44 PSIIHGWLDF-----LSSKRLVTLARGLSPAELFGGKRTDFLOQNLNPAK 92
Db 545 PESSQDFWLDHKFPLGTIIYVNTSTKSHSTSQEPLHPLNFWLNWNPQDFVAYENVNF--- 601

QY 93 SRGGPGPDYLYKNVEDDIVRSDVALDK 119
Db 602 -----DEYWKKEKEVVKQALRK 621

RESULT 10
ID Q9C5M5 PRELIMINARY; PRT; 304 AA.
AC Q9C5M5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHEITICAL 33.0 KDA PROTEIN (AT1G27000/T7N9.6).
GN T7N9.6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene T7N9.6 (GI:8778852).";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
```

```

[2]
RN SEQUENCE FROM N.A.
RP Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP Yamada K., Banh J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,
RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,
RA Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,
RA Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,
RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
RA Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene T7N9.6 (GI:8778852).";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF360143; AAK25853.1; -.
DR EMBL: AY057525; AAL09765.1; -.
DR EMBL: AY056367; AAL07253.1; -.
KW Hypothetical protein.
SQ SEQUENCE 304 AA; 33013 MW; 463DB4579AF52195 CRC64;

Query Match 13.6%; Score 84.5; DB 10; Length 304;
Best Local Similarity 27.6%; Pred. No. 0.43;
Matches 34; Conservative 20; Mismatches 44; Indels 25; Gaps 5;

QY 10 RAAGLKERTLLLDVSTK--NPVRTVNFNLSLQDPSIIH-----DGWLDFLSSKRLV 61
Db 157 RTQNLDDKVEKQIDLSKEINSQVARENISLEMDLESNLHNLITGLGKGLDTLEYKQDV 216

QY 62 TLARGSPAFLEFGGKRTDFLOQNLNPAKSRG-----GPGPDYLYKNVEDDIV 111
Db 217 TNVFMNL-LYNYFGGKSTKLPEMEQLQLPVQNRALLADVETKG-----LKNLAELF 269

QY 112 RSD 114
Db 270 KSN 272

RESULT 11
ID Q9LFY5 PRELIMINARY; PRT; 334 AA.
AC Q9LFY5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE T7N9.6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,
RA Shinn P., Altafi H., Bei Q., Chin C., Chlou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Hwang B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsy N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC T7N9 from chromosome
   I.";
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
```



RN [2]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
Khan S., Kim C., Altati H., Bei B., Chin C., Chioi J., Choi E.,  
Conn L., Conway A., Gonzalez A., Hansen N., Howling B., Koo T., Lam B.,  
Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,  
Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
Theologis A., Ecker J.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
Khan S., Kim C., Altati H., Bei B., Chin C., Chioi J., Choi E.,  
Conn L., Conway A., Gonzalez A., Hansen N., Howling B., Koo T., Lam B.,  
Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,  
Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
Theologis A., Ecker J.;  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC000348; AAF79851.1; -  
SQ SEQUENCE 334 AA; 36323 MW; 82391C5498EF3760 CRC64;

Query Match 13.6%; Score 84.5; DB 10; Length 334;  
Best Local Similarity 27.6%; Pred. No. 0.49;  
Matches 34; Conservative 20; Mismatches 44; Indels 25; Gaps 5;  
QY 10 RAAGLKEKTLILLDVSTK--NPVTVNENFLSLQDPSIIH-----DGLWDLFLSSKRLV 61  
DB 163 RIQLDDKVEKQIDLSKEINSQVISAARENSSLEMDLESNLHGLDGKLDLTLEYKQDV 222  
QY 62 TLARGLSPAFLRFGGKRTDFLQFQNLNPAKSRG-----GPGPDYYLYKNYEDDIV 111  
DB 223 TNVPMNL-LXNYFGGKSTKLPEMQLQLPVNRARNLLADVETG-----LKNLAELF 275  
QY 112 RSD 114  
DB 276 KSN 278

RESULT 12  
Q988J0  
ID Q988J0 PRELIMINARY; PRT; 756 AA.  
AC Q988J0;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
DE GLYCEROL DEHYDRATASE LARGE SUBUNIT.  
GN MLL6722.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,  
Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,

RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
Takeuchi C., Yamada M., Tabata S.;  
RT \*Complete genome structure of the nitrogen-fixing symbiotic bacterium  
Mesorhizobium loti.\*;  
RL DNA Res. 7:331-338(2000).  
DR EMBL; AP003010; BAB52960.1; -  
DR InterPro; IPR003206; Dehydratase\_IU.  
DR InterPro; IPR003208; Dehydratase\_MU.  
DR Pfam; PF02286; Dehydratase\_LU; 1.  
DR Pfam; PF02288; Dehydratase\_MU; 1.  
KW Complete proteome.  
SQ SEQUENCE 756 AA; 81889 MW; 05465FF6EB5532C9 CRC64;  
Query Match 13.3%; Score 82.5; DB 16; Length 756;  
Best Local Similarity 22.0%; Pred. No. 2.2;  
Matches 31; Conservative 25; Mismatches 52; Indels 33; Gaps 3;  
QY 2 DRRPLPVDRAA-----GLKEKTLILLDVSTKNPVTVNENFLSL 40  
DB 15 DERPLRLDKFAAEDPANGFSAPDPKPGIGIKDGRVLSLDGLVLEHDYDMIDRFIARH 74  
QY 41 QLDPSSIHDG-----WLDFLSKR--LVTLAGLSPAFLRFGGKRTDFLQFQNL 88  
DB 75 HIDEVPEAMALDSATVARKLVDNMVPRETLVRLAHGHTPAKLAEVVSQNLAEIAFAY 134  
QY 89 NPAKSRGGPGDYLYKNYEDD 109  
DB 135 SKMARKTTPGNQAHVTNAKDD 155  
RESULT 13  
Q93X83  
ID Q93X83 PRELIMINARY; PRT; 385 AA.  
AC Q93X83;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HSP70 INTERACTING PROTEIN/THIOREDOXIN CHIMERA.  
OS Vitis labrusca.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;  
OC Vitis.  
OX NCBI\_TaxID=103355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=DEVELOPING LEAF;  
RA Webb M.A., Cavaletto J.M., Klanrit P., Thompson G.A.;  
RT \*Orthologs in Arabidopsis thaliana of the Hsp-70 interacting protein,  
Hip.\*;  
RL Cell Stress and Chaperones 6:0-0(2001).  
DR EMBL; AY036906; AAK64512.1; -  
SQ SEQUENCE 385 AA; 42584 MW; 2CC963943A0D2FID CRC64;

Query Match 12.4%; Score 77; DB 10; Length 385;  
Best Local Similarity 29.5%; Pred. No. 3.9;  
Matches 28; Conservative 7; Mismatches 22; Indels 38; Gaps 3;  
QY 39 SLQIDPSIIHDGWLDFLSSKRLVTLARGLSPAFLRFGGKRTDFLQFQNLNPAK----- 92  
DB 15 SVKSDPSILHNPSLSFFKS-----YLSGLGARIPAKPKDCKG 51  
QY 93 --SRGGPGDYLYKN-----YEDDIVRSPVALD 118  
DB 52 GTSDHGEHVDKAKTNLCSEDDKFDDEDDIVESDIED 86  
RESULT 14  
Q9W3U3  
ID Q9W3U3 PRELIMINARY; PRT; 241 AA.  
AC Q9W3U3;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)





**THIS PAGE BLANK (USPTO)**

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	156.5	25.2	532	4	US-09-181-336-15	Sequence 15, Appl
2	156.5	25.2	543	2	US-08-922-170B-10	Sequence 10, Appl
3	156.5	25.2	543	2	US-09-071-739B-2	Sequence 2, Appl
4	156.5	25.2	543	4	US-09-181-336-13	Sequence 13, Appl
5	156.5	25.2	543	4	US-09-260-038B-2	Sequence 2, Appl
6	67	10.8	512	4	US-09-102-528-21	Sequence 21, Appl
7	67	10.8	584	4	US-09-102-528-12	Sequence 12, Appl
8	64.5	10.4	251	3	US-08-483-857-9	Sequence 9, Appl
9	63.5	10.2	149	2	US-08-387-942C-45	Sequence 45, Appl
10	63.5	10.2	1403	2	US-08-387-942C-3	Sequence 3, Appl
11	62.5	10.1	732	1	US-08-481-626-2	Sequence 2, Appl
12	62.5	10.1	732	4	US-08-989-299-4	Sequence 4, Appl
13	62.5	10.1	1306	4	US-08-989-299-7	Sequence 7, Appl
14	61.5	9.9	95	3	US-08-946-329A-78	Sequence 78, Appl
15	61	9.8	352	2	US-08-751-474-2	Sequence 2, Appl
16	61	9.8	649	4	US-09-188-930-305	Sequence 305, App
17	60.5	9.7	597	1	US-08-374-155A-12	Sequence 12, Appl
18	60.5	9.7	597	2	US-08-785-396-12	Sequence 12, Appl
19	60.5	9.7	1313	4	US-08-989-299-9	Sequence 9, Appl
20	60	9.7	427	4	US-09-690-265-1	Sequence 1, Appl
21	60	9.7	1015	3	US-08-463-210-9	Sequence 1, Appl
22	60	9.7	1015	4	US-09-124-900-3	Sequence 3, Appl
23	60	9.7	1016	2	US-07-743-357-2	Sequence 2, Appl
24	59.5	9.6	424	2	US-08-484-993B-18	Sequence 18, Appl
25	59.5	9.6	424	2	US-08-484-158B-18	Sequence 18, Appl
26	59.5	9.6	424	2	US-08-484-596A-18	Sequence 18, Appl
27	59.5	9.6	424	2	US-08-480-150A-18	Sequence 18, Appl

```
;
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein
; STREET: 2940 Birchtree lane
; CITY: Silver Spring
; STATE: Maryland
; COUNTRY: United States of America
; ZIP: 20906
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; OPERATING SYSTEM: Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/922,170B
; FILING DATE: 2 SEP 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedmam, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 10:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-922-170B-10
;
; Query Match 25.2%; Score 156.5; DB 2; Length 543;
; Best Local Similarity 37.2%; Pred. No. 8.7e-11;
; Matches 42; Conservative 20; Mismatches 44; Indels 7; Gaps 3;
;
; QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHD-GWLDFLSKR 59
; | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
; Db 22 GPLSPGALPRPA--QAQDVVDLDFETQEPHLVSPSFLSVTIDANLATDPRFLILGSPK 79
; | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
; QY 60 LVTIARGLSPAFLRFGKRTDFLQFQNLNRPAPKSRGGPGDPYLYKNVEDDIVR 112
; | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
; Db 80 LRTLARGLSPAYLRFGGTKTDFLIF----DPKKESTFEERSYMQSQVQNQDICK 128
;
; RESULT 3
; US-09-071-739B-2
; Sequence 2, Application US/09071739B
; Patent No. 6177545
; GENERAL INFORMATION:
; APPLICANT: Iris Pecker et al.
; TITLE OF INVENTION: HEPARANASE SPECIFIC MOLECULAR PROBES
; TITLE OF INVENTION: AND THEIR USE IN RESEARCH AND MEDICAL
; TITLE OF INVENTION: APPLICATIONS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 20001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
```

```
;
; OPERATING SYSTEM: Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,739B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/922,180
; FILING DATE: September 2, 1997
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedmam, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 2:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-09-071-739B-2
;
; Query Match 25.2%; Score 156.5; DB 4; Length 543;
; Best Local Similarity 37.2%; Pred. No. 8.7e-11;
; Matches 42; Conservative 20; Mismatches 44; Indels 7; Gaps 3;
;
; QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHD-GWLDFLSKR 59
; | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
; Db 22 GPLSPGALPRPA--QAQDVVDLDFETQEPHLVSPSFLSVTIDANLATDPRFLILGSPK 79
; | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
; QY 60 LVTIARGLSPAFLRFGKRTDFLQFQNLNRPAPKSRGGPGDPYLYKNVEDDIVR 112
; | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
; Db 80 LRTLARGLSPAYLRFGGTKTDFLIF----DPKKESTFEERSYMQSQVQNQDICK 128
;
; RESULT 4
; US-09-181-336-13
; Sequence 13, Application US/09181336A
; Patent No. 6242238
; GENERAL INFORMATION:
; APPLICANT: FREEMAN, Craig Geoffrey
; APPLICANT: HAMDORE, Brenton James
; APPLICANT: HULETT, Mark Darren
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN
; TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR
; FILE REFERENCE: 032505-004
; CURRENT APPLICATION NUMBER: US/09/181,336A
; CURRENT FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: AU PP0062
; EARLIER FILING DATE: 1997-10-28
; EARLIER APPLICATION NUMBER: AU PP0812
; EARLIER FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 13
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Homo sapiens
;
; US-09-181-336-13
;
; Query Match 25.2%; Score 156.5; DB 4; Length 543;
; Best Local Similarity 37.2%; Pred. No. 8.7e-11;
; Matches 42; Conservative 20; Mismatches 44; Indels 7; Gaps 3;
;
; QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQLDPSIIHD-GWLDFLSKR 59
```

```
Db 22 GPLSPGALPRPA--QAQDVVDLDEFTQEPHLVSPSLSVTIDANLATDPRFLILGSPK 79
QY 60 LVTLAGLSPAFRLFGGKRTDFLQFQNLNPAKSRGGPGDYLYLKNVEDD1VR 112
Db 80 LRTLARGLSPAYLRFGGTKTDFLIF----DPKKESTEERSYQSQVYNQDICK 128

RESULT 5
US-09-260-038B-2
; Sequence 2, Application US/09260038B
; Patent No. 6348344
; GENERAL INFORMATION:
; APPLICANT: Maty Ayal-Hershkovitz et al.
; TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR
; EXPRESSING RECOMBINANT HEPARANASE
; AND METHODS OF PURIFYING SAME
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 2001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; an ASCII file
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/260,038B
; FILING DATE: 02-Mar-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/071,618
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 09/071,739
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 08/922,180
; FILING DATE: September 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedmam, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/16
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-5625553
; TELEFAX: 972-3-5625554
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-260-038B-2

Query Match 25.2%; Score 156.5; DB 4; Length 543;
Best Local Similarity 37.2%; Pred. No. 8.7e-11;
Matches 42; Conservative 20; Mismatches 44; Indels 7; Gaps 3;

QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNEFLSLQLDPSIHD-GWLDLFLSKR 59
Db 22 GPLSPGALPRPA--QAQDVVDLDEFTQEPHLVSPSLSVTIDANLATDPRFLILGSPK 79
QY 60 LVTLAGLSPAFRLFGGKRTDFLQFQNLNPAKSRGGPGDYLYLKNVEDD1VR 112
Db 80 LRTLARGLSPAYLRFGGTKTDFLIF----DPKKESTEERSYQSQVYNQDICK 128
```

```
RESULT 6
US-09-102-528-21
; Sequence 21, Application US/09102528
; Patent No. 6207883
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA Sequences
; NUMBER OF SEQUENCES: 32
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/102,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/GB96/03191
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 512 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..512
; OTHER INFORMATION: /note= "Figure 9, sequence of
; At41a"
US-09-102-528-21

Query Match 10.8%; Score 67; DB 4; Length 512;
Best Local Similarity 27.2%; Pred. No. 6.8;
Matches 25; Conservative 9; Mismatches 24; Indels 34; Gaps 5;

QY 38 LSLQLDPSI-----IH-----DGWLDLFLSKRLVTLARGLSPA 70
Db 76 LNFQIDPPVQKLYVDIHGAVWDFRHHYRGTPRRHLLTGWSKFPVNSKKLIA---GDSVW 132
QY 71 FLREGGKRTDFLQFQNLNPAKSRGGPGDY 102
Db 133 FMR--KSADEMYIGVRRTPISSSDG-GSSYY 160

RESULT 7
US-09-102-528-12
; Sequence 12, Application US/09102528
; Patent No. 6207883
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA Sequences
; NUMBER OF SEQUENCES: 32
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/102,528
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/GB96/03191
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 584 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```









;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Eli Lilly and Company  
;; STREET: Lilly Corporate Center  
;; CITY: Indianapolis  
;; STATE: Indiana  
;; COUNTRY: U.S.  
;; ZIP: 46285  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/751,474  
;; FILING DATE:  
;; CLASSIFICATION: 536  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Webster, Thomas D.  
;; REGISTRATION NUMBER: 39,872  
;; REFERENCE/DOCKET NUMBER: X-9902  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 317-276-3334  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 352 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-751-474-2

Query Match 9.88; Score 61; DB 2; Length 352;  
Best Local Similarity 32.3%; Pred. No. 22;  
Matches 20; Conservative 8; Mismatches 28; Indels 6; Gaps 3;  
  
QY 32 TVNENFLSLQDPSIIHDGW-LDFLSKRLV----TLARGLSPAFLRFG-GKRTDFLOFQ 85  
Db 11 TVGHVTLNLLMPKFIEDGWEVHYIGDKRGIEHIOELKSLDVTTFHSIATGKLRFPFSWQ 70  
  
QY 86 NL 87  
Db 71 NM 72

Search completed: July 30, 2002, 08:15:05  
Job time: 287 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:14:00 ; Search time 141.47 Seconds  
(without alignments)  
387.074 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_42\_534  
Perfect score: 2577  
Sequence: 1 GDRRLPVDRAAGLKEKTLI.....PVTMGFFVKNYNALACRYR 493

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2577	100.0	534	22 AAB85216	Heparanase-like pr
2	2564	99.5	492	22 AAB84664	Amino acid sequenc
3	2538	98.5	592	22 AAU07424	Human heparanase-1
4	2538	98.5	592	22 AAB85215	Heparanase-like pr
5	2538	98.5	592	22 AAU97632	Human heparanase,
6	2528	98.1	592	22 AAB81062	Human heparanase-2
7	2277	88.4	538	22 AAU97633	Human heparanase,
8	2270	88.1	439	22 AAU07423	Human heparanase-1
9	2270	88.1	480	22 AAU07418	Novel human extrac
10	2270	88.1	480	22 AAB85217	Heparanase-like pr
11	2270	88.1	480	22 AAU97634	Human heparanase,

12	1112	43.2	214	22 AAM99905	Human excretory re
13	1112	43.2	214	22 AAM43704	Human bladder anti
14	1106	42.9	262	22 AAM24147	Human EST encoded
15	944.5	36.7	535	21 AAB08851	A murine heparanas
16	930.5	36.1	543	20 AAY02345	A human heparanase
17	930.5	36.1	543	21 AAB08849	Amino acid sequenc
18	930.5	36.1	543	21 AAY57590	Human heparanase.
19	930.5	36.1	543	21 AAY52990	Human heparanase p
20	930.5	36.1	543	22 AAY97635	Human heparanase p
21	930.5	36.1	543	22 AAY02346	A human heparanase
22	930.5	36.1	592	21 AAB08850	Amino acid sequenc
23	927.5	36.0	543	22 AAB88361	Human membrane or
24	926.5	36.0	530	20 AAY34173	Human pre-prohepar
25	926.5	36.0	543	20 AAY17082	Human heparanase e
26	926.5	36.0	543	22 AAB86206	Human heparanase i
27	926.5	36.0	588	20 AAY30124	Human heparanase i
28	883.5	34.3	532	20 AAY17083	A human protein wi
29	805	31.2	380	20 AAY17085	Seq ID No: 15 of W
30	779	30.2	380	20 AAY17084	Rat heparanase enz
31	778	30.2	156	22 AAG65963	Mouse heparanase e
32	468.5	18.2	488	22 AAB31459	Human heparanase-1
33	440.5	17.1	488	22 AAB31472	Amino acid sequenc
34	439.5	17.1	488	22 AAB31470	Amino acid sequenc
35	422.5	16.4	488	22 AAB31471	Amino acid sequenc
36	278	10.8	137	22 AAG65961	Human heparanase-1
37	278	10.8	159	22 AAG65964	Human heparanase-1
38	270	10.5	112	22 AAU07425	Human heparanase-1
39	247	9.6	256	21 AAG13479	Arabidopsis thalia
40	203	7.9	118	22 AAG65962	Human heparanase-1
41	171	6.6	935	22 AAB69219	Drosophila melanog
42	162	6.3	115	22 AAB85218	Mouse heparanase-1
43	144	5.6	38	20 AAY34186	Human pre-prohepar
44	127	4.9	24	22 AAB85220	Heparanase-like pr
45	127	4.9	25	22 AAB85221	Heparanase-like pr

ALIGNMENTS

RESULT 1  
AAB85216  
ID AAB85216 standard; Protein: 534 AA.  
AC AAB85216;  
XX  
XX  
XX 07-SEP-2001 (first entry)  
XX  
XX Heparanase-like protein Hpa2 splice variant #2.  
XX  
XX Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;  
XX KW cytosolic; neuroprotective; cerebroprotective; immunosuppressive;  
XX KW antipsoriatic; nootropic; antiinflammatory; antiarthritic; antiasthmatic;  
XX KW antidiabetic; antiarteriosclerotic; vulnery.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200146392-A2.  
XX  
XX PD 28-JUN-2001.  
XX  
XX PF 21-DEC-2000; 2000WO-GB04963.  
XX  
XX PR 22-DEC-1999; 99GB-0030392.  
XX PR 07-APR-2000; 2000GB-0008713.  
XX  
XX PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
XX  
XX PI McKenzie EA, Stamps AC, Terrett JA, Tyson KL;  
XX  
XX DR WPI; 2001-418056/44.  
XX DR N-PSDB; AAB22672.  
XX  
XX PT Novel homologs of heparanase, present in three splice variants, useful

PT for identifying agents that modulate heparanase, useful in the  
PT treatment and/or prophylaxis of abnormal levels of heparanase -  
XX  
XX  
XX Claim 1; Fig 2; 97pp; English.

XX The invention provides a homologue to heparanase which is present in  
CC three splice variants. The heparanase homologue polypeptide is useful in  
CC the treatment of a human or non-human animal or for use in diagnosis.  
CC Vectors comprising the heparanase homologue polynucleotides are useful in  
CC the transformation or transfection of a prokaryotic or eukaryotic host.  
CC The modulators of the polypeptide are useful in the manufacture of a  
CC medicament for the treatment and/or prophylaxis of a condition/disease  
CC associated with abnormal levels of the heparanase homologue, including  
CC cancer, central nervous system (CNS) and neurodegenerative diseases.  
CC cardiovascular diseases such as restenosis following angioplasty and  
CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,  
CC allografts, inflammatory diseases, arthritis, vascular restenosis,  
CC tumour growth and progression, asthma, Alzheimer's disease, diabetic  
CC retinopathy, wound healing and inflammation. The polypeptide is also  
CC useful in diagnosis and research. the present sequence represents the  
CC amino acid sequence of the mid-sized splice variant of the heparanase-  
CC like protein Hpa2 of the invention.

XX Sequence 534 AA;

Query Match 100.0%; Score 2577; DB 22; Length 534;  
Best Local Similarity 100.0%; Pred. No. 3.8e-252;  
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSKRL 60  
DB 42 GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSKRL 101  
QY 61 VTLARGLSPAFIRFGGKRTDFLQFONLRNPAKSRGGPGDPYILKNYEDDIVRSDVALDKQ 120  
DB 102 VTLARGLSPAFIRFGGKRTDFLQFONLRNPAKSRGGPGDPYILKNYEDDIVRSDVALDKQ 161  
QY 121 KGCKIAQHPDVMLEQREKAQMHLVLLKEQFSNTYSNLILTEPNNYRTMHGRAVNGSQL 180  
DB 162 KGCKIAQHPDVMLEQREKAQMHLVLLKEQFSNTYSNLILTEPNNYRTMHGRAVNGSQL 221  
QY 181 KGCKIAQHPDVMLEQREKAQMHLVLLKEQFSNTYSNLILTEPNNYRTMHGRAVNGSQL 240  
DB 222 KGCKIAQHPDVMLEQREKAQMHLVLLKEQFSNTYSNLILTEPNNYRTMHGRAVNGSQL 281  
QY 241 GRVVKMDFLKRRLDITLSDQIRKIQKVNTYTPGKKIWLCEVYVTSAGGTNNLSDSYAA 300  
DB 282 GRVVKMDFLKRRLDITLSDQIRKIQKVNTYTPGKKIWLCEVYVTSAGGTNNLSDSYAA 341  
QY 301 GFLWLNTLQMLANQGVVIRHSFDDHGVNHLVDQFNPLPDYWLSLLYKRLIGPKVLAV 360  
DB 342 GFLWLNTLQMLANQGVVIRHSFDDHGVNHLVDQFNPLPDYWLSLLYKRLIGPKVLAV 401  
QY 361 HVAGLQKRPGRVIRDKLRIYAHCTNNHNNHYVRGSTITFLINLHRSRKKTKLAGTLRD 420  
DB 402 HVAGLQKRPGRVIRDKLRIYAHCTNNHNNHYVRGSTITFLINLHRSRKKTKLAGTLRD 461  
QY 421 KLHVQYLLOPYQEGGLSKSVOLNGOPLVMVDGTLPELKPRLRAGRTLVIPTVTMGFF 480  
DB 462 KLHVQYLLOPYQEGGLSKSVOLNGOPLVMVDGTLPELKPRLRAGRTLVIPTVTMGFF 521  
QY 481 VVKNNALACRYR 493  
DB 522 VVKNNALACRYR 534  
RESULT 2  
ID AAB84664  
XX AAB84664 standard; Protein; 492 AA.  
AC AAB84664;  
XX

DT 05-SEP-2001 (first entry)  
DE Amino acid sequence of human heparanase-like polypeptide.  
XX Human; heparanase-like polypeptide; gene therapy; cancer; angiogenesis;  
KW trauma; autoimmune disease; skin disease; cardiovascular disease;  
KW nervous system disease; inflammation; arthritis; genitalia;  
KW male fertility; erectile dysfunction.  
XX Homo sapiens.  
OS  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 407 /note= "unspecified residue encoded by KCA"  
FT  
XX WO200148161-A2.  
XX  
PD 05-JUL-2001.  
XX  
PF 18-DEC-2000; 2000WO-EP12909.  
XX  
PR 23-DEC-1999; 99EP-0125831.  
XX  
PA (SCHD ) SCHERING AG.  
XX  
PI Siemeister G, Weiss B;  
XX  
DR WPI: 2001-418259/44.  
DR N-PSDB; AAH28347.  
XX  
PT Human Heparanase-like polynucleotide encoding polypeptides useful for  
PT modulating expression of the polypeptide and for treating cancer, -  
PT cancer metastasis, aberrant angiogenesis by gene therapy technique -  
XX  
PS Claim 9; Page 30; 30pp; English.  
XX  
CC The present sequence represents a human heparanase-like polypeptide.  
CC Heparanase-like polynucleotides are useful as a source of probes,  
CC primers and antisense molecules, and in gene therapy. Heparanase-like  
CC polynucleotides and polypeptides are useful for treating several  
CC disorders e.g., cancer, cancer metastasis. The oligonucleotides are  
CC also useful as diagnostic markers for the diagnosis of disorder such  
CC as cancer, cancer metastasis and aberrant angiogenesis. They may also  
CC act as diagnostic markers for diagnosis of disorder such as cancer,  
CC cancer metastasis and aberrant angiogenesis. The heparanase polypeptides  
CC and polynucleotides are also useful for treating trauma, autoimmune  
CC diseases, skin diseases, cardiovascular diseases, nervous system  
CC diseases, and inflammation including arthritis. Since the polynucleotide  
CC is preferentially expressed in male genitalia, modulation of its  
CC expression and/or activity may be used for medical intervention in male  
CC genitalia function that is male fertility control, erectile dysfunction.  
XX  
SQ Sequence 492 AA;

Query Match 99.5%; Score 2564; DB 22; Length 492;  
Best Local Similarity 99.6%; Pred. No. 6.9e-251;  
Matches 490; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 2 DRRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSKRLV 61  
DB 1 DRRPLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFLSLQDPSIIHDGWLDFLSKRLV 60  
QY 62 TLARGLSPAFIRFGGKRTDFLQFONLRNPAKSRGGPGDPYILKNYEDDIVRSDVALDKQK 121  
DB 61 TLARGLSPAFIRFGGKRTDFLQFONLRNPAKSRGGPGDPYILKNYEDDIVRSDVALDKQK 120  
QY 122 GKCKIAQHPDVMLEQREKAQMHLVLLKEQFSNTYSNLILTEPNNYRTMHGRAVNGSQL 181  
DB 121 GKCKIAQHPDVMLEQREKAQMHLVLLKEQFSNTYSNLILTEPNNYRTMHGRAVNGSQL 180  
QY 182 KDYIOLKSLQPIRIYSRASLYGPNIGRPNKRVIALLDGFMKVAGSTVDVATWQHCHYIDG 241  
DB 182 KDYIOLKSLQPIRIYSRASLYGPNIGRPNKRVIALLDGFMKVAGSTVDVATWQHCHYIDG 241

Db 181 kdiylkllqpiriyraslygnigrprknvialldgfmkvagstvdavtwqhcyyidg 240  
Qy 242 RVKVMDFKRLDLTSDQIRKIQKVNTYTPGKKIWLEGVWTSAGGTNNLSDSYAAG 301  
Db 241 rrvkvmdfkrlldtldsqirkiqkvntytpgkklwlegvvttsaggttnlstdsyaag 300  
Qy 302 FLWNLTLGMLANQIDVIRHSFFDHGYNHLYDQFNPLPDYWSLLYKRLIGPKVLAVH 361  
Db 301 flwntlgmlanqgidvvrhsffdhgynhlvdqfnplpdysllykrligpkvlavh 360  
Qy 362 VAGLQRPGRVIRDKLRIYAHCTNHHNHNVRGSIITLFIINLHRSRKKIKLAGTLRDK 421  
Db 361 vaglqrprgrvirdkrlriyahctnhhnhvrgsiltfiinlhrxrkklkgtlrdk 420  
Qy 422 LVHQLLPYGOEGLKSKSVQLNGOPLVMVDGTLPELKPRLRAGRTLVIPTVTMGFFV 481  
Db 421 lvhqllypygqeglksksvqlngqplvmvddgtlpeikprlragrtlviptvtmgffv 480  
Qy 482 VKNVNALACRYR 493  
Db 481 vknvnalacryr 492  
RESULT 3  
AAU07424  
ID AAU07424 standard; Protein: 592 AA.  
XX  
AC  
XX  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Human heparanase-like protein splice variant #1.  
XX  
KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
KW antiproliferative; cardiac; vasotropic; cerebroprotective; neurotropic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; cardiovascular disease;  
KW nervous system disorder; Alzheimer's disease; cerebral ischaemia; infection;  
KW wound healing; food additive; heparanase.  
XX Homo sapiens.  
XX  
XX WO200179253-A1.  
XX  
XX 25-OCT-2001.  
XX  
XX 11-APR-2001; 2001WO-US11643.  
XX PF  
XX  
XX 18-APR-2000; 2000US-198123P.  
XX PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX PA  
XX  
XX Fiscella M, Shi Y, Ebner R, Ruben SM;  
XX  
XX WPI; 2001-611720/70.  
XX DR N-PSDB; AAS13848.  
XX  
XX New nucleic acids encoding extracellular matrix polypeptides, for  
XX diagnosing, treating, preventing or ameliorating human disorders and  
XX disorders -  
XX  
XX Disclosure; Page 14; 308pp; English.  
XX  
XX The invention relates to novel isolated polynucleotides (I) encoding  
XX extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by  
XX (I) are used to prevent, treat or ameliorate a medical condition in e.g.  
XX humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They  
XX are also used in diagnosing a pathological condition or susceptibility to  
XX a pathological condition. The antibodies to the polypeptides can also be  
XX used in alleviating symptoms associated with the disorders and in

CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities. The present sequence represents the amino acid sequence  
CC of human heparanase-like protein, splice variant #1.  
XX  
XX Sequence 592 AA;  
SQ  
Query Match 98.5%; Score 2538; DB 22; Length 592;  
Best Local Similarity 89.5%; Pred. No. 4e-248;  
Matches 493; Conservative 0; Mismatches 0; Indels 58; Gaps 1;  
Qy 1 GDRRLPVDRAGLKEKTLILLDVSTKNPRTVNENFSLQLDPSIIHDGWLDFLSSKRL 60  
Db 42 gdrpripvdraaglkektlilldvstknprtvtvnfnfslqldpsilhdgwldfllsskrl 101  
Qy 61 VTLARGLSPAFLRFGGKRTDFLOFONLRNPAKSRGCGPDYVYKKNVEDDIVRSDVALDKQ 120  
Db 102 vtlarglspaflrfggkrtdflofqnlrnpaksr99gpdyyknyeddivrsvdvaldkq 161  
Qy 121 KGCKTAQHPDVMLELQREKAAQMHLVLKKEQFSNTYSNLILT----- 162  
Db 162 kgckiaqhpdvmllelqrekaqmhlvlkkgfsntysnliltarsldklynfadcsglhl 221  
Qy 163 -----EPNNYRTMHGRAVNSQLGK 182  
Db 222 ifalnrlrnpnnswnssalsllkysaskkyniswelgnepnnyrtmhgravnsglqgk 281  
Qy 183 DYIQLKSLLOPIRYSRASLYGPNIGRPNKVNIALLDGFMKVAGSTVDVAVTWHQCYIDGR 242  
Db 282 dyiqklsllqpirlyrsaslygpnigrprknvialldgfmkvagstvdavtwqhcyyidgr 341  
Qy 243 VVKVMDFLKTRLLDLSQIRKIQKVNTYTPGKKIWLEGVWTSAGGTNNLSDSYAAGF 302  
Db 342 vvkvmdfktrlltdltdsqirkiqkvntytpgkklwlegvvttsaggttnlstdsyaagf 401  
Qy 303 LWLNTLGLMANQIDVIRHSFFDHGYNHLYDQFNPLPDYWSLLYKRLIGPKVLAVHV 362  
Db 402 lwlntlgmlanqgidvvrhsffdhgynhlvdqfnplpdysllykrligpkvlavhv 461  
Qy 363 AGLQRKPRGRVIRDKLRIYAHCTNHHNHNVRGSIITLFIINLHRSRKKIKLAGTLRDKL 422  
Db 462 aglqrkprgrvirdkrlriyahctnhhnhvrgsiltfiinlhrxrkklkgtlrdk 521  
Qy 423 VHQLLPYGOEGLKSKSVQLNGOPLVMVDGTLPELKPRLRAGRTLVIPTVTMGFFV 482  
Db 522 vhqyllpygqeglksksvqlngqplvmvddgtlpeikprlragrtlviptvtmgffv 581  
Qy 483 KNVNVALACRYR 493  
Db 582 knvnalacryr 592  
RESULT 4  
AAB85215  
ID AAB85215 standard; Protein: 592 AA.  
XX  
AC AAB85215;  
XX  
DT 07-SEP-2001 (first entry)  
XX  
DE Heparanase-like protein Hpa2 splice variant #1.

```
XX Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;
KW cytosolic; neuroprotective; cerebroprotective; immunosuppressive;
KW antiparasitic; neurotropic; antiinflammatory; antiarthritic; antiasthmatic;
KW antidiabetic; antiarteriosclerotic; vulnerary.
XX
OS Homo sapiens.
FH
FT Key Location/Qualifiers
FT Misc-difference 237
FT /label= unknown
FT /note= "encoded by ANC"
XX
XX WO200146392-A2.
XX
XX 28-JUN-2001.
XX
XX 21-DEC-2000; 2000WO-CB04963.
XX
XX 22-DEC-1999; 99GB-0030392.
XX 07-APR-2000; 2000GB-0008713.
XX
XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.
XX
XX McKenzie EA, Stamps AC, Terrett JA, Tyson KL;
XX WPI; 2001-418056/44.
XX
XX Novel homologs of heparanase, present in three splice variants, useful
XX for identifying agents that modulate heparanase, useful in the
XX treatment and/or prophylaxis of abnormal levels of heparanase
XX
XX Claim 1; Fig 1; 97pp; English.
XX
XX The invention provides a homologue to heparanase which is present in
XX three splice variants. The heparanase homologue polypeptide is useful in
XX the treatment of a human or non-human animal or for use in diagnosis.
XX Vectors comprising the heparanase homologue polynucleotides are useful in
XX the transformation or transfection of a prokaryotic or eukaryotic host.
XX The modulators of the polypeptide are useful in the manufacture of a
XX medicament for the treatment and/or prophylaxis of a condition/disease
XX associated with abnormal levels of the heparanase homologue, including
XX cancer, central nervous system (CNS) and neurodegenerative diseases,
XX cardiovascular diseases such as restenosis following angioplasty and
XX allografts, autoimmune diseases, psoriasis, lupus erythematosus,
XX atherosclerosis, inflammatory diseases, arthritis, vascular restenosis,
XX tumour growth and progression, asthma, Alzheimer's disease, diabetic
XX retinopathy, wound healing and inflammation. The polypeptide is also
XX useful in diagnosis and research. The present sequence represents the
XX amino acid sequence of the largest splice variant of the heparanase-like
XX protein Hpa2 of the invention.
XX
XX Sequence 592 AA;
XX
XX Query Match 98.5%; Score 2538; DB 22; Length 592;
XX Best Local Similarity 89.5%; Pred. No. 4e-248;
XX Matches 493; Conservative 0; Mismatches 0; Indels 58; Gaps 1;
XX
XX 1 GDRPLPVDRAAGLUKETLLILDVSTKNPVRTVNENFLSLQLOPSIHDWDLFSKRL 60
XX
XX 42 gdrplpvdraaglukektllildvstknpvrtnenflslqldpsiihdgwlfdlskrl 101
XX
XX 61 VTLARGSLPAPLRFGGKRTDFLOFNLRNPAKSRGGPGDYILKNYEDDILVRSDVALDKQ 120
XX
XX 102 vtlargslpafirfggkrtdflofnlrnpaksrvggpgdyilknayeddivrsdvaldkq 161
XX
XX 121 KGCKIAQHPDVMLELQREKAQMHLVLLKEQFSNTYSNLIIT----- 162
XX
XX 162 kgckiaqhpdvmlqlrekaqmhlvllkeqfsntysnliitarsldklynfadcsghl 221
XX
XX 163 -----EPNNYRTMHGRAVNSQLGK 182
```

```
Db 222 ifalnairnnpnswxssalsllkysaskkyniswelgnepnnnyrtmhgravnsglqk 281
Qy 183 DYIOLSKLQPIRIYGRASLYGNIGRPKNVIALIDGFMKVGASTVDATWQHCVIDGR 242
Db 282 dyiqklsllqpiriyraslygnigrknvialldgfmkvagstvdavtwqncylodr 341
Qy 243 VVKVMDFLKTRLLDTSQIRKIQKVNTYTPGKKIWLQGVVTTSGAGTNNLSDSYAAGF 302
Db 342 vkvmdfllktrlltdsqirklqkvvntycpgkkilwlegvvttsaggttnladsyaagf 401
Qy 303 LWLNTLGLMLANOGIDVVIRHSFFDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHV 362
Db 402 lwntlgmlanogidvvirhsffdhgynhlvdqnfplpdywlslllykrligpkvlavhv 461
Qy 363 AGLQKRPGRVIRDKLRIYAHCTNHHNHNHVRGSIITLFIINLHRSRKKIKLAGTIRDKL 422
Db 462 aglqrkprgrvirdklriyahctnnhnhnyvrgsiflfinhhrsrrkkiklagtirdkl 521
Qy 423 VHQYLLQPYGQGLSKSVOLNGQPLVMVDGTLPELKPRLRAGRTLVIPTVMTGFFVV 482
Db 522 vhyllqpygqglksksvqlngqplvmvdgtlpeikprlragrtlvipvtmgffvv 581
Qy 483 KVNVALACRYR 493
Db 582 knvnalacryr 592
XX
XX RESULT 5
XX AAY97632
XX ID AAY97632 standard; Protein: 592 AA.
XX AC AAY97632;
XX XX
XX DT 20-APR-2001 (first entry)
XX DE Human heparanase, hnhp1, protein sequence.
XX KW Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
XX KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
XX KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
XX KW gene therapy; human.
XX OS Homo sapiens.
XX PN WO200100643-A2.
XX PD 04-JAN-2001.
XX PF 19-JUN-2000; 2000WO-IL00358.
XX PR 25-JUN-1999; 99US-0140801.
XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX PI Pecker I, Michal I, Itzhaki H;
XX DR WPI; 2001-137930/14.
XX DR N-PSDB; AAA91097.
XX
XX New polynucleotides and polypeptides that are distantly homologous to
XX heparanase, useful in wound healing, as well as in gene therapy
XX protocols for angiogenesis, restenosis, atherosclerosis, or
XX inflammation -
XX
XX Claim 10; Fig 1; 67pp; English.
XX
XX This sequence represents a heparanase of the invention.
XX The heparanase DNA and protein sequences are useful in wound healing,
XX angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary
XX diseases, neurodegenerative diseases (such as Scrape, Alzheimer's
XX disease, and Creutzfeldt-Jakob disease) or viral infections. The
XX heparanase coding sequence is particularly useful in gene therapy.
```



XX	SQ	Sequence	592 AA;	
Query Match				
Best Local Similarity 98.5%; Score 2538; DB 22; Length 592;				
Matches 493; Conservative 0; Mismatches 0; Indels 58; Gaps 1;				
QY	1	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNEFLSLQLDPSIIHDGWLDFSSKRL	60	
Db	42	gdrpdpvdraaglkectllildvstknprtvenfnlsldpsihdgwldfsskrl	101	
QY	61	VTLAGLSPAFIRFGGKRTDFLOFNLRNPAKSRGGPGDPYILKNYEDDIVRSDVALDKQ	120	
Db	102	vtlarginpafirfggkrtdfiqfnlrnpaksrpggdpdyilknayeddivrsvdvaldkq	161	
QY	121	KGCKIAQHPDVMLELQREAAQMHLVLLKEQFSNTYSNLILT-----	162	
Db	162	kgckiaqhpdvmlqlrekaaqmhlvllkeqfsntysnliltarsldklynfadcsghl	221	
QY	163	-----EPNNYRTMHGRAVNGSQLGK	182	
Db	222	ifalnallrnpnnswnssalsllkysaskkyniswelgnepnnyrtmhgravngsqlgk	281	
QY	183	DYIQLKSLLOPIRIYSRASLYGNIPRKNVIALLDGFMKVAGSTVDVATWQHCVIDGR	242	
Db	282	dyiqklsllqpiriysraslygniprknvialldgfmkvagstvdavtwqhcyidgr	341	
QY	243	VVKVMDFLKTRLLDTSQIRKIQKVNTYTPGKKIWLKGVVTTSGAGTNNLSDSYAAGF	302	
Db	342	vvkvmfdlktllldtsdqirkiqkvntytpgkklwlkvgvttsgaggtnnlsdsyaagf	401	
QY	303	LWLTGLMLANQIDVIRHSPFDHGYNHLVDQNFPLPDYILSLYKRLIGPKVLAVHV	362	
Db	402	lwltglmlangidvvrhspfdhgynhlvdqfnplpdyilsllykrligpkvlavhv	461	
QY	363	AGLQKRPGRVIRDKLRIYACHTNNHNNYVSGSTITLFIINHRSKKIKIAGTLURDKL	422	
Db	462	aglqkrpgrvirdklriyachtnhnnhnyvrgstlfiinlhrskkikiagtlurdkl	521	
QY	423	VHQYLQPVQGEGLSKSVQNLQGPVLMVDDGTLPELKPRPRAGRTLVIPTMGFFVY	482	
Db	522	vhqyllqpyqgeglksksvqlngqplvmvddgtlpeikprpragrtlviipvtmgffv	581	
QY	483	KNVNALACRYR 493		
Db	582	knvnalacryr 592		
RESULT 6				
ID	AAB81062	standard; Protein; 592 AA.		
XX	AAB81062;			
DT	20-JUN-2001	(first entry)		
DE	Human Heparanase-2	amino acid sequence.		
KW	Heparanase 2; human; endoglucuronidase; heparan sulphate; metastasis;			
KW	neovangiogenesis; vaccine; autoimmune disorder; blood coagulation;			
KW	cancer; diabetes; ischaemia; sepsis; stroke; cardiovascular; thrombosis.			
OS	Homo sapiens.			
PH	Key	Location/Qualifiers		
FT	Region	156..169		
FT	Region	/label= Immunogenic_epitope		
FT	Region	249..262		
FT	Region	/label= Immunogenic_epitope		
FT	Region	505..518		
FT	Region	/label= Immunogenic_epitope		
XX				

PN	WO200121814-A1.			
XX				
PD	29-MAR-2001.			
XX				
PF	11-SEP-2000; 2000WO-EP08837.			
XX				
PR	23-SEP-1999; 99EP-0118805.			
PR	07-JUL-2000; 2000EP-0114649.			
XX	(MERE ) MERCK PATENT GMBH.			
PA				
XX				
PI	Dueker K, Sirrenberg C;			
XX				
DR	WPI; 2001-308089/32.			
DR	N-PSDB; AAP86101.			
XX				
PT	New heparanase-2 polypeptide useful in diagnosing (the susceptibility			
PT	of a subject to) and as vaccines against e.g. autoimmune disorders,			
PT	cardiovascular disease, cancer, diabetes, ischemia, sepsis, stroke, or			
PT	thrombosis			
XX				
PS	Claim 1; Page 42-43; 46pp; English.			
XX				
CC	This invention relates to a human heparanase-2 protein and the cDNA			
CC	sequence encoding it. Heparanase-2 is a member of the endoglucuronidase			
CC	family of polypeptides and it degrades heparan sulphate proteoglycans			
CC	HSPGs (ubiquitous macromolecules of cell surfaces, basement membranes and			
CC	the extracellular matrix). HSPGs support the vascular endothelium and			
CC	stabilise the structure of the capillary wall. Heparanases may be			
CC	associated with neovangiogenesis and metastasis related to malignant			
CC	tumour formation. Heparanase-2 polynucleotides and proteins are useful as			
CC	vaccines for inducing an immunological response against autoimmune			
CC	disorders, blood coagulation disorders, cancer, diabetes, ischaemia,			
CC	sepsis, stroke, cardiovascular diseases, or thrombosis, as well as in			
CC	diagnosing (the susceptibility of a subject to) these diseases.			
CC	Heparanase-2 fragments may be used as immunogens to produce antibodies			
CC	immunospecific to the polypeptides, and to identify membrane bound			
CC	soluble receptors, agonists or antagonists that compete with the binding			
CC	of the polypeptide to the receptors. An antibody specific for			
CC	heparanase-2 can be used in the diagnosis of the above diseases and in			
CC	isolating or identifying clones expressing heparanase-2. The present			
CC	sequence represents heparanase-2. Three regions of heparanase-2 with high			
CC	immunogenicity (immunogenic epitopes) can be used to raise antibodies			
CC	against heparanase-2.			
XX				
SQ	Sequence 592 AA;			
Query Match				
Best Local Similarity 98.1%; Score 2528; DB 22; Length 592;				
Matches 491; Conservative 1; Mismatches 1; Indels 58; Gaps 1;				
QY	1	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNEFLSLQLDPSIIHDGWLDFSSKRL	60	
Db	42	gdrpdpvdraaglkectllildvstknprtvenfnlsldpsihdgwldfsskrl	101	
QY	61	VTLAGLSPAFIRFGGKRTDFLOFNLRNPAKSRGGPGDPYILKNYEDDIVRSDVALDKQ	120	
Db	102	vtlarginpafirfggkrtdfiqfnlrnpaksrpggdpdyilknayeddivrsvdvaldkq	161	
QY	121	KGCKIAQHPDVMLELQREAAQMHLVLLKEQFSNTYSNLILT-----	162	
Db	162	kgckiaqhpdvmlvlqrekaaqmhlvllkeqfsntysnliltarsldklynfadcsghl	221	
QY	163	-----EPNNYRTMHGRAVNGSQLGK	182	
Db	222	ifalnallrnpnnswnssalsllkysaskkyniswelgnepnnyrtmhgravngsqlgk	281	
QY	183	DYIQLKSLLOPIRIYSRASLYGNIPRKNVIALLDGFMKVAGSTVDVATWQHCVIDGR	242	
Db	282	dyiqklsllqpiriysraslygniprknvialldgfmkvagstvdavtwqhcyidgr	341	
QY	243	VVKVMDFLKTRLLDTSQIRKIQKVNTYTPGKKIWLKGVVTTSGAGTNNLSDSYAAGF	302	



XX DR WPT; 2001-611720/70.

XX PT New nucleic acids encoding extracellular matrix polypeptides, for

XX PT diagnosing, treating, preventing or ameliorating human disorders and

XX PT disease, such as, autoimmune, hyperproliferative or cardiovascular

XX PT disorders -

XX PS Disclosure; Page 13-14; 308pp; English.

XX CC The invention relates to novel isolated polynucleotides (I) encoding

XX CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by

XX CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.

XX CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They

XX CC are also used in diagnosing a pathological condition or susceptibility to

XX CC a pathological condition. The antibodies to the polypeptides can also be

XX CC used in alleviating symptoms associated with the disorders and in

XX CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked

XX CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated

XX CC include autoimmune diseases e.g. rheumatoid arthritis,

XX CC hyperproliferative disorders e.g. neoplasms of the breast or liver,

XX CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders

XX CC e.g. cerebral ischaemia, angiodenesis, nervous system disorders e.g.

XX CC Alzheimer's disease, infections caused by bacteria, viruses and fungi

XX CC be used to aid wound healing and epithelial cell proliferation, to

XX CC prevent skin aging due to sunburn, to maintain organs before

XX CC transplantation, for supporting cell culture of primary tissues, to

XX CC regenerate tissues and in chemotaxis. The polypeptides can also be used

XX CC as a food additive or preservative to increase or decrease storage

XX CC capabilities. The present sequence represents the amino acid sequence

XX CC of human heparanase-like protein.

XX SQ Sequence 439 AA;

Query Match 88.13; Score 2270; DB 22; Length 439;

Best Local Similarity 89.08; Pred. No. 3.8e-221;

Matches 439; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 GDRPLPVDRAAGLKETLILLDYSTKNPVTNENFLSLQDPSIHDGWLDFLSSKRL 60

Db 1 gdrplpvdraaglkettillldystknpvtntnenflslqdpshldgwlflsskrl 60

QY 61 VTLARGSPAFRFGKRTDFLOFNLRNPAKSRGGPGDYIYLNKNEYDDIVRSVDALDKQ 120

Db 61 vtlargspafirfggkrtdflofnlrnpaksggpgdyiynkneyddivrsvdaldkq 120

QY 121 KGCKIAQHPDVMLEQREKAAQMHLLVLLKKEQFSNTYSNLIITPENNYTMTMGRVNGSQL 180

Db 109 -----epnnyrmtmgravngsq 126

QY 181 GKDYIQLKSLLOPIRIYSRASLYGNTPGRKPNVIALLDGFMKVGASTVDATWQHCVIID 240

Db 127 gkdyiqklsllqpiyrsaslygnpgrprknvialldgfmkvagstvdattwqhcyid 186

QY 241 GRVVKVMDFLKTRLLDLSQIRIKQVNTYTPGKTIWLEGVVTTSAGGTNNLSDSYAA 300

Db 187 grvvkvmfdktrlldtsdqirkiqkvntypgkkiwlegvvttsaggtnnlstdsyaa 246

QY 301 GFLMNLTLGLMANGIDVVIHRSFDDHGNHLVDQNFNPLDPYWLSLLYKRLIGPKVLAV 360

Db 247 gflwntlglmangidvvihrsfddhgnhlvdqnfnpdpwywlsliylkrligpkvlav 306

QY 361 HVAGLQKRPGRVIRDLRIYAHTNHNHNHNYVRGSLTFLINLHRSKKIKLAGTLRD 420

Db 307 hvaglkprprgvrirldkriyachtntnnyvrsgstflfihnlsrkkiklagtlrd 366

QY 421 KLVHQYLLQPTGQBLKSKSVQLMGQPLVMVDGTLPELKPRLRAGRRLVIPPVTMGFF 480

Db 367 klvhyllqpygqgklsksvqlmgqplvmvddgtlpeklprlragrtlvipvtmgff 426

QY 481 VVKNNALACRYR 493

Db 427 vvknnalacryr 439

RESULT 9

AAU07418

ID AAU07418 standard; Protein; 480 AA.

XX AC AAU07418;

XX DT 18-DEC-2001 (first entry)

XX DE Novel human extracellular matrix (ECM) protein #1.

XX KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

XX KW antiproliferative; cardiac; vasotropic; cerebroprotective; neurotropic;

XX KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;

XX KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;

XX KW hyperproliferative disorder; neoplasm; cardiovascular disorder;

XX KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;

XX KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;

XX KW wound healing; food additive.

XX OS Homo sapiens.

XX WO200179253-A1.

XX PD 25-OCT-2001.

XX PF 11-APR-2001; 2001WO-US11643.

XX PR 18-APR-2000; 2000US-198123P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Fiscella M, Shi Y, Ebner R, Ruben SM;

XX WPI; 2001-611720/70.

XX N-PSDB; AAS13843.

XX PT New nucleic acids encoding extracellular matrix polypeptides, for

XX PT diagnosing, treating, preventing or ameliorating human disorders and

XX PT disease, such as, autoimmune, hyperproliferative or cardiovascular

XX PT disorders -

XX PS Claim 1; Page 292-293; 308pp; English.

XX CC The invention relates to novel isolated polynucleotides (I) encoding

XX CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by

XX CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.

XX CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They

XX CC are also used in diagnosing a pathological condition or susceptibility to

XX CC a pathological condition. The antibodies to the polypeptides can also be

XX CC used in alleviating symptoms associated with the disorders and in

XX CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked

XX CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated

XX CC include autoimmune diseases e.g. rheumatoid arthritis,

XX CC hyperproliferative disorders e.g. neoplasms of the breast or liver,

XX CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders

XX CC e.g. cerebral ischaemia, angiodenesis, nervous system disorders e.g.

XX CC Alzheimer's disease, infections caused by bacteria, viruses and fungi

XX CC and ocular disorders e.g. corneal infection. The polypeptides can also

XX CC be used to aid wound healing and epithelial cell proliferation, to

XX CC prevent skin aging due to sunburn, to maintain organs before

XX CC transplantation, for supporting cell culture of primary tissues, to

XX CC regenerate tissues and in chemotaxis. The polypeptides can also be used

XX CC as a food additive or preservative to increase or decrease storage

XX CC capabilities. The present sequence represents the amino acid sequence

XX CC of novel human extracellular matrix (ECM) protein #1.

XX SQ Sequence 480 AA;

Query Match	88.1%;	Score 2270;	DB 22;	Length 480;
Best Local Similarity	89.0%;	Pred. No. 4.4e-221;		
Matches 439;	Conservative 0;	Mismatches 0;	Indels 54;	Gaps 1;
Qy	1	GDRPLPVDRAAGLKEKTLILLDVTSTKNPVTYNNENFLSLQLDPSIITHDGLDFLSKRL	60	
Db	42	gdrplpvdraaglikektlilldvstknprtvtvnenflslqlpsiihdgwlfdlsskrl	101	
Qy	61	VTLAGLSPAFIRFGGKRTDFLQFNLRNPAKSRGGPGPDYLYLKNYEDDIVRSDVALDKQ	120	
Db	102	vtlaglspafirfggkrtdfllqfnlrnpaksrpgpdylylknedy-----	149	
Qy	121	KGCKIAQHPDVMLELQREKAAQMHVLVLLKEQFSNTYSNLIILTEPNNTYRTHMGRVNGSQL	180	
Db	150	-----epnnyrtmhgravnsgql	167	
Qy	181	GKDYIQLKSLLOPIRIYSRASLYGNIGRPRKNVTALLDGMKVGAGSTVDVAVTWHQCYID	240	
Db	168	gkdyiqkllqpiiriyaslygnigrprknvialldgfmkvagstvdavtwghcyid	227	
Qy	241	GRVVKVMDFLKTRLLDLSDDQIRKIQVNTYTPGKKIWLGVVTTTSAGGTNNLSDSYAA	300	
Db	228	grvvkvmfdlkrlltdltsdqirkiqkvntytpgkikiwlegvvtttsaggtnnlsdsyaa	287	
Qy	301	GFLWNTLGLMLANQIDVYIRHSFFDHGYNHLVDQNFNPLDPDYWLILLYKRLIGPKVLAV	360	
Db	288	gflwntlgmlanqgidvyrhshffdhgynhlvdqnfnpdpdywllslykrligpkvlav	347	
Qy	361	HVAGLQRPGRVIRDKLRIYAHCTNNHNNHYVRSSTLTFTINLHRSRKKIKLAGTLRD	420	
Db	348	hvaglqrkprgrvirdklriyahctnnhnnhyvrgsstlftinlhrsrrkkiklagtlrd	407	
Qy	421	KLHVQYLLQPYGQEGLSKSVQNGOPLVMVDDGTLPELKPRPLRAGRTLVIPTVMTGFF	480	
Db	408	klvhqyllqpyggeglksksvqnggplvmvddgtlpeikprplragrtlvipptvmgff	467	
Qy	481	VKKNVNALACRYR	493	
Db	468	vkknvnalacryr	480	
RESULT	10			
AAB85217		standard; Protein; 480 AA.		
ID	AAB85217			
AC	AAB85217;			
XX				
DT	07-SEP-2001 (first entry)			
DE	Heparanase-like protein Hpa2 splice variant #3.			
DE				
KW	Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;			
KW	cytostatic; neuroprotective; cerebroprotective; immunosuppressive;			
KW	antipsoriatic; neutropic; antiinflammatory; antiarthritic; antiasthmatic;			
KW	antidiabetic; antiarteriosclerotic; vulneryary.			
XX				
OS	Homo sapiens.			
XX				
PN	WO200146392-A2.			
XX				
PD	28-JUN-2001.			
XX				
PF	21-DEC-2000; 2000WO-GB04963.			
XX				
PR	22-DEC-1999; 99GB-0030392.			
PR	07-APR-2000; 2000GB-0008713.			
XX				
XX	(OXFO-) OXFORD GLYCOSCIENCES UK LTD.			
XX				
PI	McKenzie EA, Stamps AC, Terrett JA, Tyson KL;			
XX				
XX	WPI; 2001-418056/44.			
DR	N-PSDB; AAH22673.			
DR				

XX	Novel homologs of heparanase, present in three splice variants, useful for identifying agents that modulate heparanase, useful in the treatment and/or prophylaxis of abnormal levels of heparanase -
PT	
PT	
XX	Claim 1; Fig 3; 97pp; English.
XX	
CC	The invention provides a homologue to heparanase which is present in three splice variants. The heparanase homologue polypeptide is useful in the treatment of a human or non-human animal or for use in diagnosis.
CC	Vectors comprising the heparanase homologue polynucleotides are useful in the transformation or transfection of a prokaryotic or eukaryotic host.
CC	The modulators of the polypeptide are useful in the manufacture of a medicament for the treatment and/or prophylaxis of a condition/disease associated with abnormal levels of the heparanase homologue, including cancer, central nervous system (CNS) and neurodegenerative diseases.
CC	cardiovascular diseases such as restenosis following angioplasty and atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus, allografts, inflammatory diseases, arthritis, vascular restenosis, tumour growth and progression, asthma, Alzheimer's disease, diabetic retinopathy, wound healing and inflammation. The polypeptide is also useful in diagnosis and research. The present sequence represents the amino acid sequence of the smallest splice variant of the heparanase-like protein Hpa2 of the invention.
CC	
XX	Sequence 480 AA;
SQ	
	Query Match 88.1%; Score 2270; DB 22; Length 480;
	Best Local Similarity 89.0%; Pred. No. 4.4e-221;
	Matches 439; Conservative 0; Mismatches 0; Indels 54; Gaps 1;
Qy	1 GDRPLPVDRAAGLKEKTLILLDVTSTKNPVTYNNENFLSLQLDPSIITHDGLDFLSKRL 60
Db	42 gdrplpvdraaglikektlilldvstknprtvtvnenflslqlpsiihdgwlfdlsskrl 101
Qy	61 VTLAGLSPAFIRFGGKRTDFLQFNLRNPAKSRGGPGPDYLYLKNYEDDIVRSDVALDKQ 120
Db	102 vtlaglsfafirfggkrtdflqfnlrnpaksrpgpdylylknedy----- 149
Qy	121 KGCKIAQHPDVMLELQREKAAQMHVLVLLKEQFSNTYSNLIILTEPNNTYRTHMGRVNGSQL 180
Db	150 -----epnnyrtmhgravnsgql 167
Qy	181 GKDYIQLKSLLOPIRIYSRASLYGNIGRPRKNVTALLDGMKVGAGSTVDVAVTWHQCYID 240
Db	168 gkdyiqkllqpiiriyaslygnigrprknvialldgfmkvagstvdavtwghcyid 227
Qy	241 GRVVKVMDFLKTRLLDLSDDQIRKIQVNTYTPGKKIWLGVVTTTSAGGTNNLSDSYAA 300
Db	228 grvvkvmfdlkrlltdltsdqirkiqkvntytpgkikiwlegvvtttsaggtnnlsdsyaa 287
Qy	301 GFLWNTLGLMLANQIDVYIRHSFFDHGYNHLVDQNFNPLDPDYWLILLYKRLIGPKVLAV 360
Db	288 gflwntlgmlanqgidvyrhshffdhgynhlvdqnfnpdpdywllslykrligpkvlav 347
Qy	361 HVAGLQRPGRVIRDKLRIYAHCTNNHNNHYVRSSTLTFTINLHRSRKKIKLAGTLRD 420
Db	348 hvaglqrkprgrvirdklriyahctnnhnnhyvrgsstlftinlhrsrrkkiklagtlrd 407
Qy	421 KLHVQYLLQPYGQEGLSKSVQNGOPLVMVDDGTLPELKPRPLRAGRTLVIPTVMTGFF 480
Db	408 klvhqyllqpyggeglksksvqnggplvmvddgtlpeikprplragrtlvipptvmgff 467
Qy	481 VKKNVNALACRYR 493
Db	468 vkknvnalacryr 480
RESULT	11
AA97634	
ID	AA97634 standard; Protein; 480 AA.
XX	

AC AAY97634;  
XX 20-APR-2001 (first entry)  
DT Human heparanase, hnhp1 pn5 form, protein sequence.  
XX Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scarape;  
DE atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;  
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;  
KW gene therapy; human.  
XX Homo sapiens.  
OS  
XX WO200100643-A2.  
PN 04-JAN-2001.  
XX 19-JUN-2000; 2000WO-IL00358.  
XX 25-JUN-1999; 99US-0140801.  
PR (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
XX Pecker I, Michal I, Itzhaki H;  
XX WPI; 2001-137930/14.  
DR N-PSDB; AAA91099.  
XX New polynucleotides and polypeptides that are distantly homologous to  
PT heparanase, useful in wound healing, as well as in gene therapy  
PT protocols for angiogenesis, restenosis, atherosclerosis, or  
PT inflammation -  
XX  
XX Claim 10; Page 63; 67pp; English.  
XX  
XX This sequence represents a heparanase of the invention.  
CC The heparanase DNA and protein sequences are useful in wound healing,  
CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary  
CC diseases, neurodegenerative diseases (such as Scarape, Alzheimer's  
CC disease, and Creutzfeldt-Jakob disease) or viral infections. The  
CC heparanase coding sequence is particularly useful in gene therapy.  
XX  
SQ Sequence 480 AA;  
  
Query Match 88.1%; Score 2270; DB 22; Length 480;  
Best Local Similarity 89.0%; Pred. No. 4.4e-221;  
Matches 439; Conservative 0; Mismatches 0; Indels 54; Gaps 1;  
  
QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVTNENFSLQLDPSIIHDGWLDFLSKRL 60  
Db 42 gdrplpvdraaglkektllldvstknpvtvnfnslqlpslihdgwldfisskrl 101  
QY 61 VTLAGLSPALFREGKRTDFLQFNLRNPAKSRGGPGPDYLLKNYEDDIYRSDVALDKQ 120  
Db 102 vtlaglsparifrggkrtdfqlgnlrnpaksrpggpgpdyllknayed----- 149  
QY 121 KGCKIAQHPDVMLEQREAAQMLVLLKEQFSNTYSNLIITEPNNTMHRGAVNGSQL 180  
Db 150 -----epnnrytmgravnsgql 167  
QY 181 GKDYIQLKSLLOPTRIYRSASLYGPNIGRPRKNVIALLDGPMKVGSTVDVATWQHICYD 240  
Db 168 gkdyiqikslilqiriyraslygpnigrprknvialldgfmkvagstvdatwqhcyid 227  
QY 241 GRVVKVMDFLKRLDLLDLSQIRKIQKVVNTYTPGKKIWLGVVTTGAGTNNLSDSYAA 300  
Db 228 grvvkvmdfiktrlldtsdqirkiqkvntytpgkkiwlgvvttsgagtnnlsdsyaa 287  
QY 301 GFLWNTLGLMANGIDWIRHSFFDHGYNHLVDQNFNPLPDYWLSLLYKRLIGPKVLAV 360  
Db 288 gflwntlglmandgadvirhsffdhgynhlvdqnfnpdpdywlslllykrligpkvlav 347

QY 361 HVAGLQKRPGRVIRDKLRIYAHCTNHHNNHNYVRSITLFIINLHRSRKKIKIAGTLRD 420  
Db 348 hvaglrkrprgrvirdklriyahctnhnhnyvrgsitlfiinlhrsrkkikiagtlrd 407  
QY 421 KLVHQYLLQPYGQSEGLKSKSVQNLNGOPLVMVYDDGCTLPDLKPRPLRAGRTLVIPTVMGFF 480  
Db 408 klvhyllqpygggglksksvqlngqplvmvddgtlplkprprlragrtlviptvmgff 467  
QY 481 VVKNVNALACRYR 493  
Db 468 vvkvnalacryr 480  
  
RESULT 12  
AAM99905  
ID AAM99905 standard; Protein; 214 AA.  
XX  
AC AAM99905;  
XX  
DT 07-JAN-2002 (first entry)  
XX  
DE Human excretory related polypeptide SEQ ID NO 642.  
XX  
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antischizoid; antianemic; antithrombotic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antinflammatory;  
KW antiallergic; antidiabetic; antidiabetic; antidiabetic; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;  
KW excretory system.  
XX  
OS Homo sapiens.  
XX  
PN WO200155313-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01323.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 30-JUN-2000; 2000US-0211535.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.

PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 14-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 13-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 01-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 03-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-465569/50.  
XX N-PSDB; AA198878.  
DR  
DR Isolated nucleic acid molecule encoding excretory system antigen is  
DR used in preventing, treating or ameliorating a medical condition -  
XX  
XX Claim 11; SEQ ID NO 642; 574pp + Sequence Listing; English.  
XX  
XX The invention relates to novel excretory system related human  
XX polynucleotides (AA198567-AA199503) and the encoded proteins  
XX (AA199594-AA199593) useful for preventing, treating or ameliorating  
XX medical conditions e.g. by protein or gene therapy, especially  
XX disorders related to the excretory system. The genes are isolated  
XX from a range of human tissues disclosed in the specification. The  
XX nucleic acids, proteins, antibodies and (ant)agonists are useful in  
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone  
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
XX colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
XX and parasitic infections.  
XX Note: The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 214 AA;

Query Match 43.2%; Score 1112; DB 22; Length 214;  
Best Local Similarity 100.0%; Pred. No. 3.5e-104;  
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 284 VTTSAGGTNNLSDSYAAGFLWTLGMLANQGDIVRIHSFDDHYNHLYDONFPLPDY 343  
|||||

Db	5	vtsaggttnlstdsyaagflwntlglmlanggidvvirhsffdhgynhlwqdnfnplpdy	64
QY	344	WLSLLYKRLIGPKVLAVHVAGLQRKPRGVRIRDLKRIYAHCTNHHNNHYVRGSTITFII	403
Db	65	WLSLLYKRLIGPKVLAVHVAGLQRKPRGVRIRDLKRIYAHCTNHHNNHYVRGSTITFII	124
QY	404	NLHRSRKIKLAGTLGRDLKLVHGYLLQPYGQBELSKSVQLNGQPLVMVDDGTLPELKRPP	463
Db	125	nlhrrsfkiklagtlrldklvhqyllqpygqeglkksvqlngqplvmvddgtlpeklprp	184
QY	464	LRAGRTLVPPTMGFFWKNVNALACRYR	493
Db	185	lragrtlivpvtmgffwknvnalacryr	214
<hr/>			
RESULT 13			
AA043704	ID	AA043704 standard; Protein; 214 AA.	
XX	XX	AA043704;	
AC	AC	AA043704;	
XX	XX	24-OCT-2001 (first entry)	
DT	DT	Human bladder antigen, SEQ ID NO: 98.	
DE	DE		
XX	XX		
XX	XX	Human; bladder antigen; cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer; vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic; cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder; wound healing; infection; neurological disease.	
KW	KW	Homo sapiens.	
OS	OS	WO200159064-A2.	
PN	PN	16-AUG-2001.	
XX	XX	17-JAN-2001; 2001WO-US01342.	
PF	PF	31-JAN-2000; 2000US-0179065.	
XX	XX	04-FEB-2000; 2000US-0180628.	
PR	PR	24-FEB-2000; 2000US-0184664.	
PR	PR	02-MAR-2000; 2000US-0186350.	
PR	PR	16-MAR-2000; 2000US-0189874.	
PR	PR	17-MAR-2000; 2000US-0190076.	
PR	PR	18-APR-2000; 2000US-0198123.	
PR	PR	19-MAY-2000; 2000US-0205515.	
PR	PR	07-JUN-2000; 2000US-0209467.	
PR	PR	28-JUN-2000; 2000US-0214886.	
PR	PR	30-JUN-2000; 2000US-0215135.	
PR	PR	07-JUL-2000; 2000US-0216647.	
PR	PR	07-JUL-2000; 2000US-0216880.	
PR	PR	11-JUL-2000; 2000US-0217487.	
PR	PR	11-JUL-2000; 2000US-0217496.	
PR	PR	14-JUL-2000; 2000US-0218290.	
PR	PR	26-JUL-2000; 2000US-0220963.	
PR	PR	26-JUL-2000; 2000US-0220964.	
PR	PR	14-AUG-2000; 2000US-0224518.	
PR	PR	14-AUG-2000; 2000US-0224519.	
PR	PR	14-AUG-2000; 2000US-0225213.	
PR	PR	14-AUG-2000; 2000US-0225214.	
PR	PR	14-AUG-2000; 2000US-0225266.	
PR	PR	14-AUG-2000; 2000US-0225267.	
PR	PR	14-AUG-2000; 2000US-0225268.	
PR	PR	14-AUG-2000; 2000US-0225270.	
PR	PR	14-AUG-2000; 2000US-0225447.	
PR	PR	14-AUG-2000; 2000US-0225757.	
PR	PR	14-AUG-2000; 2000US-0225758.	
PR	PR	14-AUG-2000; 2000US-0225759.	
PR	PR	18-AUG-2000; 2000US-0226279.	
PR	PR	22-AUG-2000; 2000US-0226681.	
PR	PR	22-AUG-2000; 2000US-0226686.	
PR	PR	22-AUG-2000; 2000US-0227182.	
PR	PR	23-AUG-2000; 2000US-0227009.	
PR	PR	30-AUG-2000; 2000US-0228924.	
PR	PR	01-SEP-2000; 2000US-0229287.	
PR	PR	01-SEP-2000; 2000US-0229343.	
PR	PR	01-SEP-2000; 2000US-0229344.	
PR	PR	01-SEP-2000; 2000US-0229345.	
PR	PR	05-SEP-2000; 2000US-0229349.	
PR	PR	05-SEP-2000; 2000US-0229513.	
PR	PR	06-SEP-2000; 2000US-0230437.	
PR	PR	06-SEP-2000; 2000US-0230438.	
PR	PR	08-SEP-2000; 2000US-0231243.	
PR	PR	08-SEP-2000; 2000US-0231244.	
PR	PR	08-SEP-2000; 2000US-0231244.	
PR	PR	08-SEP-2000; 2000US-0231413.	
PR	PR	08-SEP-2000; 2000US-0231414.	
PR	PR	08-SEP-2000; 2000US-0232080.	
PR	PR	08-SEP-2000; 2000US-0232081.	
PR	PR	12-SEP-2000; 2000US-0231966.	
PR	PR	14-SEP-2000; 2000US-0232397.	
PR	PR	14-SEP-2000; 2000US-0232398.	
PR	PR	14-SEP-2000; 2000US-0232399.	
PR	PR	14-SEP-2000; 2000US-0232400.	
PR	PR	14-SEP-2000; 2000US-0232401.	
PR	PR	14-SEP-2000; 2000US-0232403.	
PR	PR	14-SEP-2000; 2000US-0233064.	
PR	PR	14-SEP-2000; 2000US-0233065.	
PR	PR	21-SEP-2000; 2000US-0234223.	
PR	PR	21-SEP-2000; 2000US-0234274.	
PR	PR	25-SEP-2000; 2000US-0234997.	
PR	PR		

PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249285.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI WPI; 2001-514652/56.  
DR N-PSDB; AAI64065.  
XX  
XX Forty five bladder related polynucleotides, useful in the prevention,  
PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
PT disorders and neurological diseases -  
XX  
XX Claim 11; SEQ ID NO 98; 482pp + sequence listing; English.  
PS  
XX The invention relates to forty five novel bladder related  
CC polynucleotides. The polynucleotides and the polypeptides that they  
CC encode are useful in the diagnosis, treatment and prevention of:  
CC cancer, particularly breast and ovarian cancer, and other cancers  
CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal  
CC tract, liver, lung, or urogenital system; immune disorders such as  
CC Addison's disease, allergies, autoimmune haemolytic anaemia,  
CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple  
CC sclerosis, rheumatoid arthritis and ulcerative colitis; cardiovascular  
CC disorders such as myocardial ischaemias; wound healing; neurological  
CC diseases such as cerebral anoxia and epilepsy; and infectious diseases  
CC such as viral, bacterial, fungal and parasitic infections. Numerous  
CC examples of each type of disorder are given in the specification.  
CC The polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities. The polynucleotides are  
CC useful for chromosome identification. They are also useful as probes  
CC for diagnosing or treating a disorder related to the female reproductive  
CC system, particularly breast and/or ovary cancer. The present  
CC sequence is a novel bladder antigen provided in the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 214 AA;

Query Match 43.2%; Score 1112; DB 22; Length 214;  
Best Local Similarity 100.0%; Pred. No. 3.5e-104;  
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 VTTSAGTNNLSDSYAAGFLWLTGLMLANOGIDVIRHSEFDHGYNHLVDQNFNPLPDY 343  
Db 5 vttsagtnnlsdsyaagflwltglmlanogldvirhsffdhgyhnlvdqnfnpdpdy 64  
QY 344 WLSLYKRLIGPKVLAVHVHAGLQKRPGRVIRDKLRIYAHCTNNHHNNYVRSITLFI 403  
Db 65 wlslykrligpkvlavhvahglqrkprpgrvirdklriyahctnnhhnhyvrsitlfi 124  
QY 404 NLHRSRKKIKLAGTLDKLVHQLLQPYGEGLSKSVQLNGOPLVMVDDGTLPKLPKP 463  
Db 125 nlhrrskkiklagtdklvghyllqpyggeglsksvqlngqplvmvddgtlpelkprp 184  
QY 464 LRAGRTLVIPVTMGFFVKNVNALACRYR 493  
Db 185 lragrtlvipvtcmgffvknvnalacryr 214  
RESULT 14  
AAM24147  
ID AAM24147 standard; Protein; 262 AA.  
XX  
XX AAM24147;  
XX  
XX 12-OCT-2001 (first entry)  
XX Human EST encoded protein SEQ ID NO: 1672.  
XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
KW diagnostics; forensic test; gene mapping; genetic disorder;  
KW biodiversity; gene therapy; nutrition.  
XX  
XX Homo sapiens.  
XX  
XX OS  
XX PN WO200154477-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 25-JAN-2001; 2001WO-US02687.  
XX  
XX 25-JAN-2000; 2000US-0491404.  
PR 17-JUL-2000; 2000US-0617746.  
PR 03-AUG-2000; 2000US-0631451.  
PR 15-SEP-2000; 2000US-0663870.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
PI Cao Y, Drmanac RA, Zhang J, Werhman T;  
XX  
XX WPI; 2001-476164/51.  
DR N-PSDB; AAH98806.  
XX  
XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
PT antibodies and research use -  
XX  
XX Claim 20; Page 1122-1123; 1275pp; English.  
XX  
XX The present invention provides the protein and coding sequences of novel  
CC proteins from a variety of organisms, including human, dog, cat, horse,  
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
CC from the organism of interest. They can be used in diagnostics,  
CC forensics, gene mapping, identification of mutations, to assess  
CC biodiversity and for nutritional purposes. The present sequence is a  
CC protein of the invention.  
XX  
XX Sequence 262 AA;

Query Match 42.9%; Score 1106; DB 22; Length 262;  
Best Local Similarity 98.2%; Pred. No. 1.9e-103;





---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:16:24 ; Search time 71.68 seconds  
(without alignments)  
660.882 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_42\_534  
Perfect score: 2577  
Sequence: 1 GDRRLPVDRAAGLKEKTLI.....PVTMGFFVKNVNNALACRYR 493

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2270	88.1	480	2 JC7506	heparanase protein
2	302	11.7	521	2 T45608	hypothetical prote
3	155	6.0	190	2 T01953	hypothetical prote
4	111	4.3	1260	2 T14022	reverse transcript
5	109	4.2	986	2 F98229	sarcosine oxidase
6	109	4.2	986	2 AH3056	glutamate dehydrog
7	105	4.1	1020	2 T18342	protein-tyrosine k
8	101.5	3.9	629	2 C64180	cytochrome-c oxida
9	101.5	3.9	1125	1 JH0771	adenylate cyclase
10	101	3.9	535	2 S58740	hypothetical prote
11	101	3.9	2026	1 O1BY	probable glucosyl
12	100.5	3.9	582	2 A84206	hypothetical prote
13	100.5	3.9	630	2 C71374	probable membrane
14	99.5	3.9	455	2 S67627	hypothetical prote
15	99.5	3.9	1272	2 C90593	PRP8 protein homol
16	99	3.8	2403	2 T30875	tryptophan--trna l
17	98.5	3.8	335	2 B84992	endopeptidase Clp
18	97.5	3.8	845	1 G72079	ClpC proteinase [i
19	97.5	3.8	845	2 C86545	ATP-dependent Clp
20	97.5	3.8	845	2 G81590	DNA-directed RNA P
21	97.5	3.8	1517	2 B81393	DNA polymerase III
22	96.5	3.7	969	2 D82895	probable helicase
23	96	3.7	763	2 B84432	cytoplasmic aconit
24	96	3.7	798	2 B84471	SIP3 protein - yea
25	96	3.7	1229	2 S42391	qsopa protein - Co
26	95.5	3.7	406	2 S68866	hypothetical prote
27	95.5	3.7	472	2 H90563	hypothetical prote
28	95.5	3.7	965	2 T22933	hypothetical prote
29	95	3.7	456	2 T24442	hypothetical prote

30	95	3.7	512	2 T11261	cytochrome-c oxida
31	95	3.7	557	2 D97210	extracellular neut
32	95	3.7	740	2 T51619	probable ethylene
33	94.5	3.7	460	2 G90554	p48-like (mycoplas
34	94.5	3.7	1280	2 B34087	hypothetical prote
35	94.5	3.7	1391	2 T20642	hypothetical prote
36	94.5	3.7	1397	2 E87998	protein F09C3.1 [i
37	94	3.6	625	2 G86855	glucose inhibited
38	94	3.6	902	2 D71079	hypothetical prote
39	94	3.6	3194	2 D71917	toxin-like outer m
40	93	3.6	320	2 D90478	conserved hypothet
41	93	3.6	484	2 C86264	protein kin-15 [im
42	93	3.6	488	2 I44330	protein-tyrosine k
43	93	3.6	849	2 S77217	phosphorylase (EC
44	93	3.6	894	2 T15769	hypothetical prote
45	92.5	3.6	328	2 AC1210	dTDP-D-glucose 4,6

ALIGNMENTS

RESULT 1  
JC7506  
heparanase protein 2a - human  
C:Species: Homo sapiens (man)  
C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 01-Dec-2000  
C:Accession: JC7506  
R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hircok, M.;  
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000  
A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase fami  
A:Reference number: JC7506  
A:Accession: JC7506  
A:Molecule type: mRNA  
A:Residues: 1-480 <MCK>  
A:Cross-references: GB:AF282885  
C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and th  
therapies.  
C:Genetics:  
A:Gene: hpa2a  
A:Map position: 10q23-10q24  
C:KeyWords: heparin binding; membrane bound

Db 348 HVAGLQRPGRVIRDKLRIYAHCTNHHNHNHVRGSLTFLFIINLHRSKKIKLAGTLRD 407  
QY 421 KLVHQLLPQYQEGKSKSVOLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFF 480  
Db 408 KLVHQLLPQYQEGKSKSVOLNGQPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFF 467  
QY 481 VVKNNALACRYR 493  
Db 468 VVKNNALACRYR 480

RESULT 2  
T45608  
hypothetical protein F13G24.30 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T45608  
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba  
submitted to the Protein Sequence Database, December 1999  
A:Reference number: 223009  
A:Accession: T45608  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-521 <BEV>  
A:Cross-references: EMBL:AL133421  
A:Experimental source: cultivar Columbia; BAC clone F13G24  
C:Genetics:  
A:Map position: 5  
A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3  
A:Note: F13G24.30

Query Match 11.7%; Score 302; DB 2; Length 521;  
Best Local Similarity 24.8%; Pred. No. 1.2e-15;  
Matches 139; Conservative 73; Mismatches 183; Indels 166; Gaps 27;

QY 34 NENFLSLQD--PSIITHD-----GWLDFLS---SKRLVTLA-RGLSPAFLRFGGR 78  
Db 22 DENFVCATLDWP---HDKCNVDQCPWGYSSVINMDLTPRLTKAIFKPLRIRIGSSL 78  
QY 79 TDFLQFQ--NLNRP-----KSRGGPGPDYLLKNYEDDVRSDVALDKQKGIQAHDPVM.132  
Db 79 QDQVIYDVGNLKTCPRPFKMNSG-----LFGFSKGLHMRWD-- 117  
QY 133 LELOREKAAQMHV-----LLKEQFSNTYSNLILTEPNRYTMHGRAVN---- 176  
Db 118 -ELNSFLTATGAVVTGLNALGRHKLRKANGGAWDHINTQDFLNTYVSKGYVIDSWEF 176  
QY 177 GSOL-----GKDYIQLKSLQPIRIYSRASLYGPNIGRPRKNVIALDGF- 221  
Db 177 GNELSGSGVGASVAELYGKDLVLKDVIN--KVYKNSWLHKPIELVAP-----GGFY 226  
QY 222 -----MKVAG-STVDVAVTQHCVIDGR-----VVKVMDPLKTRLLDLSQIRKIQ 266  
Db 227 EQQWYTKLEISGPSVVDVVT-HHIYNLGSNDPALVKKIMD---PSYLSQVSKTFKQYN 282  
QY 267 KVYVNTTPGKKIWLEGVWTSAG---G7NNLSDSYAAGFLWMLTGLMLANQIDVVIH 322  
Db 283 QTIQEHGP-----WASPWGESGGAYNSGGRHSDTIDFSFWLDQUGMSARHNTKYVCQ 338  
QY 323 SFDFHGYNHLVDQNFPLPDYWLSSLYKRLIGPKVLAVHVGALQRPGRVIRDKRLIY 382  
Db 339 TLVGGFYGLLEKGTFFVPNDYYSALLHRLMKGVLAVQTDG-----PPQLRYV 387  
QY 383 AHCNTNHHNHNHVRGSLTFLFIINLH-----RSRKKIKLAGTLRDKLV 423  
Db 388 AHCCKG-----RAGVTLLIINLSQSDFTVSYSNGINVLNAESRKKSLDITLKRPF 441  
QY 424 -----HOYLLQPYGEG-LKSKSVOLNGQPLVMVDDGTLPELKPRLRAGRT 469  
Db 442 WIGSKASDGYLNREETHLP--ENGVLRSKTVLNGSKLKPATGDIPLSLPVRVSN 499  
QY 470 LVIPPTVTMGFFVVKNNALAC 490

Db 500 LNVPLSMFSFIVLPNFDASAC 520

RESULT 3  
T01953  
hypothetical protein T2L5.6 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 21-Jan-2000  
C:Accession: T01953  
R:Geisel, C.; Smith, A.; Le, T.  
submitted to the EMBL Data Library, October 1998  
A:Description: The sequence of A. thaliana T2L5.  
A:Reference number: 214470  
A:Accession: T01953  
A:Status: translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-190 <GEI>  
A:Cross-references: EMBL:AF096371; NID:g3695386; PID:g3695392  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4  
A:Introns: 36/2; 69/3  
A:Note: T2L5.6  
C:Superfamily: Arabidopsis thaliana hypothetical protein T2L5.6

Query Match 6.0%; Score 155; DB 2; Length 190;  
Best Local Similarity 24.9%; Pred. No. 8.6e-05;  
Matches 49; Conservative 32; Mismatches 70; Indels 46; Gaps 6;

QY 321 RHFSEFDHGYNHLVDQNFPLPDYWLSSLYKRLIGPKVLAVHVGALQRPGRVIRDKLR 380  
Db 12 ROSLGGNGYGLLNTNFTPNPDYISALVRLQMLGKALFTTFESGTYK-----KIR 60  
QY 381 IYAHCTNHHNHNHVRGSLTFLFIINLHRSR---KKIKLAGTLRDKLVHQLLPQY----- 431  
Db 61 SYTHCARQSK-----GITVLLMLNLTNTTVVAKVELNNSF--SLRHTKHKMSYKRASSQ 112  
QY 432 ---GQEG-----LKSQSVOLNGQPLVMVDDGTLPELKPRLRAGRTLVIPT 473  
Db 113 LFGNGVGTQREYHHTAKDGNLHSQTMLLNGNALQVNSMGDLPPIEPIHINSTEPIITA 172  
QY 474 PVTMGFFVVKNNALAC 490  
Db 173 PYSIVFVHMRNVVPAC 189

RESULT 4  
T14022  
reverse transcriptase homolog - slime mold (Dictyostellium discoideum) retrotransposab  
C:Species: Dictyostellium discoideum  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 11-May-2000  
C:Accession: T14022  
R:Winckler, T.; Tschepke, C.; de Hostos, E.L.; Jendretzke, A.; Dingermann, T.  
Mol. Gen. Genet. 257, 655-661, 1998  
A:Title: Tdd-3, a transfer RNA gene-associated poly(A) retrotransposon from Dictyoste  
A:Reference number: 217858; MUID:98265925  
A:Accession: T14022  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-1260 <WLN>  
A:Cross-references: EMBL:AF002669; NID:g2558820; PID:g2558822; PIDN:AAC48324.1

Query Match 4.3%; Score 111; DB 2; Length 1260;  
Best Local Similarity 19.7%; Pred. No. 4.1;  
Matches 129; Conservative 89; Mismatches 204; Indels 232; Gaps 34;

QY 28 NPVRTVNEFNFLSLQIDPSIIHDGWL---DFL---SSKRLV-----TLARGLSP 69  
Db 94 NGIGILNHNQNKLSPIFIEGRLLIISDILIKDTTRILAIYAPAQPKRKTLASTLNK 153

QY 70 AFLRGGKRTDLOFQNL-RNPAKSRGGPGDYLYKNYEDDIVRSDVALDKQCKIAQH 128  
Db 154 HENN-----QYHNTFSNPKNLDIIAGDFNCLDNHTSND-----DQGNLTQTS 199  
QY 129 PDVM---LELOEKAQAHLVLLKEQ--FSNTYSNLILTEPNRYTMHGRA-----VNGSQ 179  
Db 200 PDEMATVIEAIRISNNLMDQLNKRPTFSRIHN---TNNLTRILERRLDRIYLNNSL 256  
QY 180 LKQDYIQLKSLQD----- 193  
Db 257 INYSQLYLRNLIPPKINDIPLSDHNFSLTFTLHNIOTNMRRWRLKSSILSMLKNID 316  
QY 194 --IRYSR-----ASLY-----GNIGRPRKNVIALID 219  
Db 317 FLNGYSRELSNHNHSISFQSLNLLNKLKQDYTEFKQKQNDYNNKANI-----KNLSLLE 372  
QY 220 -GFMKVGSTVDVATW-----QHC-----YIDGRVVK-VMDFLKTRLDLDTL 258  
Db 373 TEFKDOAFATLAINESKREBQKQELNNYCEETSLKYSARIKRHNDFTINAVKDTQ 432  
QY 259 SQDIKRIQKVWYTT-----PGKKIMLEGVVVTSAGGTNNLSDSYAGFLW 304  
Db 433 GRTINKQELIEEYKYVSNLYDYKEDDPSPHYEILENNVTTRDSTWDLNEFTSQEIL 492  
QY 305 --LNTLGLMANOGIDVIRHSFFDH-----GYN-----HLVDQNF----- 337  
Db 493 EVIKQLNPKHSPGDPGIPNLFVITHKEKLAPILASAFNDTLRNPHLISKNYKEGLIITIP 552  
QY 338 -----NLPDYWLSLYK---RLIGPKVLAVHVGLOKRP---PGRVIRDKLRIY 382  
Db 553 KKGDPKELINRRPITLANCIYKHSKLNNRPIITKVININOKGFVPGRFILHNIISI 612  
QY 383 AHCTNHHNNYVSGSTTLFIINLHRSRKKIKLAGTLRDKL--VHQYLLQPYGO-EGLKSK 439  
Db 613 NELINVCNDKRIINGIITLYI---SKLLTRSHTVQSQIITTHQHSNOYINLIMLLTK 667  
QY 440 S---VQLNQGQPLVMWDDGLP-ELKRPRLRAGRTLVIPTVMTGFFVKKVNAIA 489  
Db 668 SEARIEINGRT-----TIPFEIK-RGVKQG-----DPLSPTLFVLV-IEALA 707  
RESULT 5  
F98229 sarcosine oxidase alpha chain PA5418 [imported] - Agrobacterium tumefaciens (strain C58,  
C:Species: Agrobacterium tumefaciens  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C:Accession: F98229  
A: Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorillo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; PMID:11743194  
A:Accession: F98229  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-986 <KUR>  
A:Cross-references: GB:AE007870; PIDN:AAK9360.1; PID:g15159208; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L1574  
A:Map position: linear chromosome

Query Match 4.2%; Score 109; DB 2; Length 986;  
Best Local Similarity 18.2%; Pred. No. 4.1;  
Matches 122; Conservative 69; Mismatches 211; Indels 268; Gaps 27;  
QY 2 DRRLPVDRAAGLK-----EKTLLLDVSTKNPVRTVNEFLSLQDPSIHDGWLDF 54  
Db 362 DQRLDEARGLIEVLAGHSVNTAGRLVSSMTVGRGSGNKRKTAIDALVVSAGW--- 418  
QY 55 LSKRLVTLIARGLSPAFLRGKRTDFLOFQNLN-----PAKSRG 96  
Db 419 TFSVHLFSOSRK-----LKFDAQRFLPDIIHVQNCVSGACNGTDLDLVAITAEAAAGG 474

QY 97 PGPDYYLYKN-----YEDDIVRSDVALDKQCKIAQH 130  
Db 475 GSAFSGENARAWTGMIGAAEGAGGTGVKAFIDFQHDVCAKDIRLAVREG----- 526  
QY 131 VMLELOREKAAQHLLVLLKEQFSNTYSNLILTEPNRYTMHGRAVNGSOLGKDY--IQLK 188  
Db 527 -----MHSV---EHKRETTNGMASDQGKMSNMHGLAIASEALGRDLPKVGLT 571  
QY 189 SLLOP-----IRIYSRASLYGPNIGRPNVIALLDG-----FMKVAGS 227  
Db 572 TFRQPTPTVFTGLINHSGALFDPTRKTPMHEE-ELAAGAVFEDVGNKRAWFFPRAGE 630  
QY 228 TVDAVTWQHCYIDGRVVKVMDFLKTRLDLDTLSQIIRKIQKV-----NTYTPGKKIWLKG 282  
Db 631 DMHEAINRECKTVRTSGVGFDASTLKGIEVVGPDAAKFLNLITYNAWDTLKPGRCRY--G 688  
QY 283 VYVTSAGGTNNLSDSYAAAGFLWNTLGLMANOGIDVIRHSFFDHGYNHLYVDNFNPLPD 342  
Db 689 IMTREDGFV-----YDDG-----VVGRLAEDRFHVTITTTGGAPRVLQHMEDYLQTEFPD 737  
QY 343 --YWLSLLYKR-----LIGPK---VLAVHVGLOKRP-----PGRV--- 374  
Db 738 LNVWLTSATEQWAVIAVQGPKAREVIAPFVEGIDLSPFAFPHMAVAEGKFCGYPTRLFRV 797  
QY 375 -----IRDKLRIYAHCTNHHNNYVSGSTTLFIINLH----- 406  
Db 798 SFTGELGFEINVPADYGAAVWSAIRDTEAVGGC-----LYGTETMHLRAEKGYII 849  
QY 407 -----RSRKKIKLAGTLRDKLVHQLQPYGOEGL 436  
Db 850 VGQDGTGTTPDDAGLAWAVSKKTKDFVGIRGLKRIDLRTGRKQLV-----GL 898  
QY 437 KKSQVQLNGQPLVMWDDG---TLPELKPRL----- 464  
Db 899 KTK-----DRLTVPDEGGQIVTDPN-QPKPMTMLGHVTSAYSENGLHSIAFALVADGR 951  
QY 465 -RAGRTLVIIP 473  
Db 952 ARMGETLYIP 961  
RESULT 6  
AH3056 sarcosine oxidase alpha subunit [imported] - Agrobacterium tumefaciens (strain C58, D  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C:Accession: AH3056  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AH3056  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-986 <KUR>  
A:Cross-references: GB:AE008689; PIDN:AAI44870.1; PID:g17742518; GSPDB:GN00187  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: soxA  
A:Map position: linear chromosome  
Query Match 4.2%; Score 109; DB 2; Length 986;  
Best Local Similarity 18.2%; Pred. No. 4.1;  
Matches 122; Conservative 69; Mismatches 211; Indels 268; Gaps 27;  
QY 2 DRRLPVDRAAGLK-----EKTLLLDVSTKNPVRTVNEFLSLQDPSIHDGWLDF 54

Db 362 DQRLDEARALGIEVLGHVSIVNTAGRLRVSSMTVGRNGGSKRRIAIDALVWSAGW--- 418  
QY 55 LSKRLVLTILARGLSPAFIRFGKRTDFLQFQNLNR-----PAKSRGG 96  
Db 419 TFSVHLFSOSRCK---LKFDAANQRLPDIHVQNCVSIAGACNGTDDLVIAIAEAAAAGG 474  
QY 97 PGPDYYLKN-----YEDDIVRSVALDKQKCKIAQHPD 130  
Db 475 GSAFSGENARAWTGMIGAAEGAGGTGKAFIDFQHDVCAKDIRLAVREG----- 526  
QY 131 VMLELOREKAAQMHVLLKQFSNTYSNLILTEPNYRTMHGRAVNSQLGKDY--IQLK 188  
Db 527 -----MHSV---EHLKRTTNGMASDQCKMSNMHGLATASEALGRDLPKVGLT 571  
QY 189 SLLOP-----IRIYSRASLYGPNIGRPRKNVIALLDG-----FMKVAGS 227  
Db 572 TFPQPTVPTFTGLINHSRGALFDPTRKTPMHEE-ELAAAGAFEDVGNWKRWFPPRAGE 630  
QY 228 TVDAVTWQHCYIDGRVVKVMDFLKTRLLDLSQIRKIQKV-----NTYTPGKKIWLGG 282  
Db 631 DMHEAINRECKTVRTSGVDFDASTLGIKIEVVGPDAAKFLNLITYNADWTLKPGRCRY--G 688  
QY 283 VVTTAGTNNLSDSYAAGFLMLNTLGLANQIDVIRHSFDFHGYNHLVQNFNPDP 342  
Db 689 IMTREGFV-----YDDG-----VVGRLAEDRFHVTITGGAPRVLQHMEDYLQTEFPD 737  
QY 343 --YWLSLLYKR-----LIGPK-----VLAVHVAGLQRPK-----PGRV--- 374  
Db 738 LNVWLTSATQWAVIAVQGPKAREVIAPFVEGIDLSPEAFPHMAVAGFCGVPTRLFV 797  
QY 375 -----IRDKLRIYAHCTNHHNHNHVGRSGITLFIINLH----- 406  
Db 798 SFTGELGFEINVPADYGAAVWSAIRORTKTEAVGCC-----LYGTETWHILRAEKGYII 849  
QY 407 -----RSRKKIKLAGTLRDKLVHVOYLLQPYQOGL 436  
Db 850 VGQDTCGTPTPDAGLAWAVSKKTDVFGIRGLKRIDLTNRGRQLV-----GL 898  
QY 437 KSKSVQLNGPLVMDG-----TLPELKRPL----- 464  
Db 899 KTK-----DRLTVPDGGQIVTDPN-OPKPMTMLGHVTSAYSENGLHSHIAFALVADGR 951  
QY 465 -RAGRTLVP 473  
Db 952 ARWGETIYIP 961  
RESULT 7  
T18342  
glutamate dehydrogenase (EC 1.4.1.2) precursor - Sauroleishmania tarentolae  
C:Species: Sauroleishmania tarentolae  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-May-2000  
C:Accession: T18342  
R:Bringaard, F.; Stripecke, R.; Frech, G.C.; Freedland, S.; Turck, C.; Byrne, E.M.; Simps  
Mol. Cell. Biol. 17, 3915-3923, 1997  
A:Title: Mitochondrial glutamate dehydrogenase from Leishmania tarentolae is a guide RNA  
A:Reference number: Z18878; MUID:97342629  
A:Accession: T18342  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1020 <BRI>  
A:Cross-references: EMBL:U31177; NID:g945092; PID:g945093; PIDN:AAB62735.1  
C:Genetics:  
A:Gene: GDH  
C:Superfamily: glutamate dehydrogenase  
C:Keywords: NAD; oxidoreductase  
Query Match 4.1%; Score 105; DB 2; Length 1020;  
Best Local Similarity 25.5%; Pred. No. 8.8;  
Matches 42; Conservative 30; Mismatches 51; Indels 42; Gaps 7;  
QY 73 RFGGKRTDFLOFQNLRNPAKSRGGP-----GPDYILKNYEDDIVRSDVA-----LDKQ 120

Db 656 KLGLKESEMRKFQT-----GGPDGDLGSNEVLRSKKNVGMVDISASLHDPNGIDRE 707  
QY 121 KGCKIAQHPDVMLELOREKAAQMHVLLKQFSNTYSNLILTEPNYRTMHGRAV-NGSQ 179  
Db 708 ELARLAHRLPLREFRSRK-----LSPEGF-----LVLTDHNVKLPDGLTIVEDGSR 754  
QY 180 LKGDYIOLKSLLOPIRITYSRASLYGPNIGRPRKNVIALLDGFMKV 224  
Db 755 LRNEFHLK-----YSDADVFPVCGGRPRSVTLNVGVREFLKV 791  
RESULT 8  
C64180  
hypotheical protein H11056 - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999  
C:Accession: C64180  
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630  
A:Accession: C64180  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-629 <TIGR>  
A:Cross-references: GB:U32786; GB:U42023; NID:g1574605; PIDN:AAC22721.1; PID:g1574609  
Query Match 3.9%; Score 101.5; DB 2; Length 629;  
Best Local Similarity 19.9%; Pred. No. 8;  
Matches 54; Conservative 50; Mismatches 98; Indels 69; Gaps 9;  
QY 33 VVENFLSLQDPSIITHGWLDFLSKRLVTLARGISPAFLREGGRRTDFLOFQNLRNPAK 92  
Db 198 IEQAYLKILMDIDIFRDNLFNTVTKTKIGVGSSE-----GSKLKDSTEFINFSKNR 252  
QY 93 SRGGPGPDY-----YLNKYEDD-----IVRSDVALDKQKCKIAQHPDVMLE 134  
Db 253 ERLFLNPVYQKTEVEFEIKNYEDSGSKWKYTOVLIDLGEKILLEKDGFKYHYFN---- 308  
QY 135 LOREKAAQMHVLLKQFSNTYSNLILTEPNN--YRTMHGRAVNSQLGKDYIQLKSLQ 192  
Db 309 -----AQMTSIVKFSQDONLSKEIITYEYSHKVYRTTNAQSSIRSKIEDLYSIKNGIV 362  
QY 193 PRIYSRASLYGPNIGRPRKNVIALL-----DGMKVAGSTV-----D 230  
Db 363 SIE-----YIPQKGNAGNLIEVFYNASNKDMFMSDLMIKEKNKYFYLOKVNTLWD 415  
QY 231 AVTWOHC-----YIDGRVVKVMDFLKTRLID 256  
Db 416 DIOYNLNKEGGYIDFKNGKKEALLRIID 446  
RESULT 9  
JH0771  
protein-tyrosine kinase (EC 2.7.1.112), receptor type hyk precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Jun-1993 #sequence\_revision 10-Jun-1993 #text\_change 21-Jul-2000  
C:Accession: JH0771  
R:Horita, K.; Yagi, T.; Kohmura, N.; Tomooka, Y.; Ikawa, Y.; Alzawa, S.  
Biochem. Biophys. Res. Commun. 189, 1747-1753, 1992  
A:Title: A novel tyrosine kinase, hyk, expressed in murine embryonic stem cells.  
A:Reference number: JH0771; MUID:93129253  
A:Accession: JH0771  
A:Molecule type: mRNA  
A:Residues: 1-1125 <HOR>  
A:Cross-references: GB:D13738; NID:g220439; PIDN:BAA02883.1; PID:g220440  
C:Genetics:  
A:Gene: hyk

C:Superfamily: protein-tyrosine kinase, receptor type tie; EGF homology; fibronectin type C; Keywords: Anp; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho F;1-22/Domain: signal sequence #status predicted <Sig>  
F;23-1125/Product: protein-tyrosine kinase, receptor type hyk #status predicted <MAT>  
F;37-104/Domain: immunoglobulin homology <IM1>  
F;135-137/Region: cell attachment (R-G-D) motif  
F;212-252/Domain: EGF homology <EG1>  
F;256-299/Domain: EGF homology <EG2>  
F;303-341/Domain: EGF homology <EG3>  
F;365-427/Domain: immunoglobulin homology <IM2>  
F;448-526/Domain: fibronectin type III repeat homology <FN3A>  
F;541-625/Domain: fibronectin type III repeat homology <FN3B>  
F;638-720/Domain: fibronectin type III repeat homology <FN3C>  
F;752-773/Domain: transmembrane #status predicted <TM>  
F;823-1100/Domain: protein kinase homology <KIN>  
F;831-839/Region: protein kinase ATP-binding motif  
F;140,158,400,439,465,559,649,691/Binding site: carbohydrate (Asn) (covalent) #status F;856,873,965/Active site: Lys, Glu, Asp #status predicted

Query Match 3.9%; Score 101.5; DB 1; Length 1125;  
Best Local Similarity 21.3%; Pred. No. 19;  
Matches 80; Conservative 43; Mismatches 118; Indels 135; Gaps 19;

Qy 13 GLKEKTLILLDVSKNPVRTVNEF-----LSLQIDPSIIHD--GWLDFL 55  
Db 701 GLEPETTHVDIFAEENIGSSNPAFSHELRTLPHSPASADLGCGKADTHSPHWGWNDA 760  
Qy 56 SSKRLVTLARGLSPAFL-----RFGKRTDFLOFNLR-NPA-----KSRGG 96  
Db 761 SPCLL-----AFLLMLQLKRVNRRMAQAFQNVREPAQFNSGTLALNRKAKN 811  
Qy 97 PGPDY-----LK-NYEDDIVRSVDALDKQG----- 122  
Db 812 PDPTIYPVLDMNDIKFDQVIGEGNFQGVKARIKKDGLRMDAAIKRMKEYASKDDHRDFA 871  
Qy 123 -----CKIAQHPDMLLELOR-ERAAQMHVLLKQEFNTY-----SNLIETPN-NYRT 169  
Db 872 GELEVLCKLGHHPNIIILGACEHGVLYLAIEYAPHGNLLDFLGKSRVLETDPAFCHQ 931  
Qy 170 MHGRAVNSQL-----GKDYIQLKSLLOPIRIYSRASLYGPNIGRPKNVIALDGG 220  
Db 932 QYSTSLSSQQLHFEAADVARGMDVLSQKFTHR-DLAARNILVG-----ENYIAKID 983  
Qy 221 FMKVAGSTVDVTHQCHYIDGRVVKVDFLCTR--LDTLSQIRKKQKVNTVTPGKKI 278  
Db 984 FGLSRGQEV-----YVKKTGRLPVRWMAIESLN-----YSVYTTNSDV 1022  
Qy 279 WLEGVW---TTSAGGT 291  
Db 1023 WSYGVLLWEIVSLGGT 1038

RESULT 10  
S58740  
Cytochrome-c oxidase (EC 1.9.3.1) chain I - yeast (Hansenula wingei) mitochondrion  
C:Species: mitochondrion Hansenula wingei  
C:Date: 19-Mar-1997 #sequence\_revision 09-May-1997 #text\_change 07-Dec-1999  
C:Accession: S58740  
R:Seikito, T.; Okamoto, K.; Kitano, H.; Yoshida, K.  
Curr. Genet. 28, 39-53, 1995  
A:Title: The complete mitochondrial DNA sequence of Hansenula wingei reveals new character  
A:Reference number: S58740; MUID:96022424  
A:Accession: S58740  
A:Molecule type: DNA  
A:Residues: 1-535 <SEK>  
A:Cross-references: EMBL:D31785  
A:Note: the authors translated the codon CAA for residue 54 as Gly  
C:Genetics:  
A:Gene: cox1  
A:Genome: mitochondrion  
A:Genetic code: SGC3  
A:Introns: 241/3

C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology C; Keywords: cytochrome; copper; electron transfer; heme; iron; lipoprotein; magnes ductase; respiratory chain; transmembrane protein  
F;11-458/Domain: cytochrome-c oxidase chain I homology <CO1>  
F;63,379/Binding site: heme a iron (His) (axial ligands) #status predicted  
F;242,291,292/Binding site: copper (His) #status predicted  
F;242-246/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
F;246/Binding site: oxygen (Tyr) #status predicted  
F;320/Binding site: myristate (Lys) (covalent) #status predicted  
F;369/Binding site: magnesium (His) (shared with chain II) #status predicted  
F;377/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 3.9%; Score 101; DB 2; Length 535;  
Best Local Similarity 22.2%; Pred. No. 6.9;  
Matches 68; Conservative 46; Mismatches 102; Indels 90; Gaps 19;

Qy 194 IRIYSRASLYGPNIGRPKNVIALDGMKVGASTVDVTHQCHYIDGRVVKVMDFLKT 252  
Db 260 VSTSKRPVGE-----ISWYVAMASIA--FLGLVWSHMYIVG-----LD-ADT 302  
Qy 253 RLDTLSQIRKKIKQVNTYTPGKKIWLGVVTTTSAGTNLSDS--YAAGFLWLNTLG- 309  
Db 303 RAYFTSSTMVIAVPTGIKIFS-----WL-----ATLYGGSIRLAVPMLYIAFLFLFTIGG 353  
Qy 310 ----MLANQGDIVIRHSFFPDHGNHLV---DQNFNPLPDYV-----LSLLYKR----- 351  
Db 354 LTGVALANASLDVAFHDTYYVGVGHFYVLSMGAIFSLFAGYYSPQLGLYFNERLAQI 413  
Qy 352 ----LIGPKV--LAVHVAGLQKRPGRVIRDKLRIYAHCTNHHNNHNVVRG----- 396  
Db 414 QFWLIFGVANVIFPMPHFLGLQGMPR-----RIDPDYDAYAGWNYVSSIGSVIAII 464  
Qy 397 SITLFIINLHRSRKKIKLAGTLRDKLVHQ---YLLQP-----YGOEGLKSKSVQ--L 443  
Db 465 SLALFYIYID-----QLINGLTNNKIDNVVYSKAPDVFESNTIFANNKSASIEFL 519  
Qy 444 NGQPLV 449  
Db 520 NSPPAI 525

RESULT 11  
OYBY

adenylate cyclase (EC 4.6.1.1) - yeast (Saccharomyces cerevisiae)  
N:Alternate names: ATP pyrophosphate-lyase; protein J1401; protein YJL005W  
C:Species: Saccharomyces cerevisiae  
C:Date: 28-Dec-1987 #sequence\_revision 08-Sep-1995 #text\_change 21-Jan-2000  
C:Accession: S56776; S56775; A24776; S05828; S55183  
R:To Van, D.; Perea, J.; Jacq, C.  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56776  
A:Accession: S56776  
A:Molecule type: DNA  
A:Residues: 1-1823 <DEH>  
A:Cross-references: EMBL:Z49280; GSPDB:GN00010; MIPS:YJL005W  
R:de Haan, M.; Grivell, L.A.; Smits, P.H.M.  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56771  
A:Accession: S56775  
A:Molecule type: DNA  
A:Residues: 673-2026 <ZAG>  
A:Cross-references: EMBL:Z49280; GSPDB:GN00010; MIPS:YJL005W  
R:Kataoka, T.; Broek, D.; Wigler, M.  
Cell 43, 493-505, 1985  
A:Title: DNA sequence and characterization of the S. cerevisiae gene encoding adenyla  
A:Reference number: A24776; MUID:86079531  
A:Accession: A24776  
A:Molecule type: DNA  
A:Residues: 1-261, 'L', 263-547, 'L', 549-591, 'H', 593-708, 'I', 710-961, 'P', 963-1387, 'S', 13  
A:Cross-references: EMBL:M12057; NID:g171359; PIDN:AAA34549.1; PID:g171360  
A:Note: the authors translated the codon TTA for residue 262 as Ser. ACG for residue 2 as Leu, TCA for residue 1566 as Ala, GAG for residue 1659 as Ala, GGT for residue 1

R;Masson, P.; Lenzen, G.; Jacquemin, J.M.; Danchin, A.  
Curr. Genet. 10, 343-352, 1986  
A:Title: Yeast adenylate cyclase catalytic domain is carboxy terminal.  
A:Reference number: S05828; MUID:88165073  
A:Accession: S05828  
A:Molecule type: DNA  
A:Residues: 1042-1426, 'D', 1428-1460, 'T', 1462-1955, 'V', 1957-2008, 'M', 2014-2015, 'T', 'T', 'N', 'F', 'Y', 'K', 'W', 'L', 'R', 'T'  
A:Cross-references: EMBL:X03449; NID:93487; PIDN:CAA27175.1; PID:93488  
R;de Haan, M.; Smits, P.H.M.; Grivell, L.A.  
submitted to the EMBL Data Library, May 1995  
A:Reference number: S55183  
A:Accession: S55183  
A:Molecule type: DNA  
A:Residues: 673-2026 <DEW>  
A:Cross-references: EMBL:X87611; NID:9854567; PIDN:CAA60917.1; PID:9854568  
C:Genetics:  
A:Gene: SGD:CYR1; MIPS:YJL005W  
A:Cross-references: SGD:S0003542; MIPS:YJL005W  
A:Map position: 10L  
C:Superfamily: yeast adenylate cyclase; leucine-rich alpha-2-glycoprotein repeat homolog  
C:Keywords: CAMP biosynthesis; duplication; phosphorus-oxygen lyase; tandem repeat  
F:669-1343/Region: leucine-rich 23-residue repeats  
F:1065-1087/Domain: leucine-rich alpha-2-glycoprotein repeat homolog <LRR>  
F:1610-2026/Domain: catalytic status predicted <CAT>  
F:1664-1749/Domain: yeast adenylate cyclase catalytic domain homolog <YACC>

Query Match 3.9%; Score 101; DB 1; Length 2026;  
Best Local Similarity 20.1%; Pred. No. 50;  
Matches 120; Conservative 80; Mismatches 201; Indels 196; Gaps 30;

QY 12 AGLKE-KTLLLDVSTKPV---RVNENFLSLQDPSI-----I 47  
DB 903 AGFVELKNQLDLSNFMHYPEVINYCTNLQIDLSNKIQSLPQSTKYLVKLAKMNL 962  
QY 48 HDGWLDFSSKRLVTLARGLSPAFRFGGKRTDFQFQNL----- 87  
DB 963 SHKNLNFIDLSMDTLRLNLRYNRISSIKTASNLQNLFLTDNRISFEETLPKRLAL 1022  
QY 88 ---RNPAGSGGGPDYLYKNVEDDIDVSDVALDKQCKTAQHPDVMLE---LQREKA 140  
DB 1023 ETQENPITSIS--FKDFYKPN-----MTSLTLNR---AQLSSIPGELLTKLSFLEKLEL 1071  
QY 141 AQHVLVLLKEQFSNTYSNLILTEPNN-----YRT--MHGRAVNGSOLGKDY 184  
DB 1072 NONNLRLPQESKTLKLVLSVARNKLEYIPPELSQLSKSLTLDLHSNNIRDFVDMEN 1131  
QY 185 IQLKSLQPIRIYSRASL-----YGNIGRPRKNVIALDGFPMKVAGSTVDVATWQ 235  
DB 1132 LETSLNISNNAFGSSLENSFVHMSYSGSKLSK-----SLM--FFTAADNQFDAMWP 1183  
QY 236 ---HCYIDGRVVKVMDPLKTRLLDLSQIRKIQKVNTVTPGKKIWLGVVTTTSAGTNN 293  
DB 1184 LFNCFYN---LKVNLISYNNFSDV---SHMKLESITELYSGNKL-----TTLSGDT-- 1229  
QY 294 LSDSYAAGFLW--LNTLGLANGQIDV-----VIRHSEFDHGYNHLVDQNFNPLPDY-W 344  
DB 1230 -----VLKWSLSKTLMLNSQMLSLPAELSNLSQSVDFVGANQKYKNISYHYDWNW 1282  
QY 345 LS---LLYKRLIGP-----KVLAVHAGLQKRPGRVI 375  
DB 1283 RNNKELKYLNFSGNRRRFEIKSFISHDIDADLSLTLPQLKVLGLMDVTLNTTKVPDENV 1342  
QY 376 RDKLRIYACHTNNH-----NVV-----RGSTTLFTLNLHRSRKK----- 411  
DB 1343 NFRLLTASTINGMRGVADTLGQRIYVSSROVTFEFGNDDDECLLHDSKNQADYG 1402  
QY 412 ---IKLAGTLRDK-LVHQYLLQPYGQE-----GLKSKSVQLNGQ---PLWVDDG 454  
DB 1403 HNISRIVRDIYDKILIRQ--LERYGDETDDNKTALRFSFLQNLKKEINGMLNSVDNG 1457

RESULT 12

A84206  
hypothetical protein porA [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky  
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483  
A:Accession: A84206  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-582 <STO>  
A:Cross-references: GB:AE004437; NID:g10580081; PIDN:AAG19013.1; GSPDB:GNO0138  
C:Genetics:  
A:Gene: porA  
C:Superfamily: Halobacterium halobium 2-oxoacid ferredoxin oxidoreductase; 2-oxoacid

Query Match 3.9%; Score 100.5; DB 2; Length 582;  
Best Local Similarity 21.2%; Pred. No. 8.5;  
Matches 69; Conservative 49; Mismatches 110; Indels 97; Gaps 14;

QY 8 VDRAAGLKEKTLILLDVSTKNPVRTVNEFLSLQDPSIHHQWDLFLSSKRLVTLARGL 67  
DB 58 VDQVQSVVDRLLDILIALTE---RTVDENLDELHADSIIIVD----- 95  
QY 68 SPAFLRFGGKRTDFQFQNLNRPASRGSGGPDYLYKNYEDD---IVRSDVALDKQKC 123  
DB 96 -----GDRTEPADFE---SPAEV---TGLDPLKDLAEDAGGATMRNLVALGAV--C 139  
QY 124 KIAQHPDVMLEQREKAAQMHVLLKKEQFSNTYSNLILTEPNVYRTHGRAVNGSOLGKD 183  
DB 140 AVADFPPIENLDESLEK-----RFGSGEQIIT---ENNKQ-----AARLGAE 177  
QY 184 YI--QLKSLQPIRIYSRASLYGNIGRPRKNVIALDGFPMKVAGSTVDVATWQHCYIDG 241  
DB 178 YVAEEFDVTLPELETDEDYVLLNGDEAIGMAIAAGCRFYAGYPTTAT----- 229  
QY 242 RVKVMDFLTKRL-----LDTLSDQIRKIQKVNTVTPGKKIWLGVVTTTSAGTNNLS 295  
DB 230 ---DVMEYLTGRIFQFGHVVQAEDELAALNLAALGAARAGAR---SWTATSGPGCIDLMS 282  
QY 296 DSYAAGFLWNLTLGLNANOQIDVVI 320  
DB 283 E-----TFGLVATSETPLVI 297

RESULT 13  
C71374  
probable glucose inhibited division protein A (gida) - syphilis spirochete  
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 26-Aug-1999  
C:Accession: C71374  
R;Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G  
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998  
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A:Reference number: A71250; MUID:98332770  
A:Accession: C71374  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-630 <COL>  
A:Cross-references: GB:AE001189; GB:AE000520; NID:g3322293; PIDN:AAG65038.1; PID:g332  
A:Experimental source: strain Nichols  
C:Genetics:  
A:Gene: TP0044  
C:Superfamily: gida protein





Tue Jul 30 10:12:06 2002

```

| :|||: | | : ||
Db 934 SGFKDLIFTDEQYDDEDTREHRSIAITTKMPRNQFIQKIVEARQNNDLDALVDYFTNL 993
QY 191 LQPIRIYSRA---SLYGNIGRPRKNVIA-----LL 218
Db 994 IAPFRISNETLSFIRTONWGSPRKNGCAHREDIADQTAEOFTILHRLSNLVTNKEKNIR 1053
QY 219 DGFMKVAGSTVDATWQHXYIDGRVYKVMDFLKTRELL--DTLSQIRKIQKVVVYTPGK 276
Db 1054 DHIREWLNETVFAIRTSH-----LKTSEEFKNEFLAKTSSERIELIKKYDYIP-- 1104
QY 277 KIWLEGVVTTSAGTNNLS-----DSYAAGF--LWLNLTGLMLANQIGIDVVIRH 322
Db 1105 ----EG-YTLNINYTHNITSKYKYDNOVERKENKLSDFEILYIITKGDKSSQ-----RK 1153
QY 323 SPFDHG---YNHLVDQNFNPL-----PDYW 344
Db 1154 LILNYGQRIYFNHPKANQLNPLFIKRDFLKYKDEYW 1189
```

Search completed: July 30, 2002, 08:16:27  
Job time: 344 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:31:27 ; Search time 36.14 Seconds  
(without alignments)  
528.189 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_42\_534  
Perfect score: 2577  
Sequence: 1 GDRRLPVDRAAGLKEKTLI.....PVTMGFFVKNVNALACRYR 493

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101.5	3.9	629	1 T3MH_HAEIN	P71366 haemophilus
2	101	3.9	2026	1 CYAA_YEAST	P08678 saccharomyc
3	100.5	3.9	630	1 GDA_TREPA	O83084 treponema p
4	98.5	3.8	335	1 SYW_BUCAI	P57602 buchnera ap
5	97.5	3.8	845	1 CLPC_CHLPN	Q928a6 chlamydia p
6	97.5	3.8	1517	1 RPOC_CAMJE	O9pi30 campylobact
7	96.5	3.7	969	1 DP3A_UREPA	Q9p974 ureaplasma
8	96	3.7	1229	1 SIP3_YEAST	P38717 saccharomyc
9	94.5	3.7	682	1 SNK_RAT	Q9r012 rattus norv
10	94.5	3.7	685	1 SNK_HUMAN	Q9nyy3 homo sapien
11	94	3.6	625	1 GIDA_IACLA	Q9cej4 lactococcus
12	93.5	3.6	947	1 LKTA_PASSP	P55123 pasteurella
13	93	3.6	488	1 K115_CAEEL	P34891 caenorhabdi
14	93	3.6	849	1 PHSG_SYNT3	P73511 synechocyst
15	92.5	3.6	682	1 SNK_MOUSE	P53351 mus musculu
16	92.5	3.6	690	1 HELS_METTH	O26901 methanobact
17	92.5	3.6	1616	1 VITI_CAEEL	P55155 caenorhabdi
18	92.5	3.6	2241	1 TEGU_HCMVA	P16785 human cytom
19	92	3.6	461	1 GPI3_YEAST	P32363 saccharomyc
20	92	3.6	534	1 COX1_KLULA	P20386 klyveromyc
21	92	3.6	535	1 COX1_HANWI	P48868 hansenula w
22	92	3.6	898	1 ACOC_CUCMA	P49608 cucurbita m
23	92	3.6	4367	1 DVHC_NEUCR	P45443 neurospora
24	91.5	3.6	1259	1 LIN1_HUMAN	P08547 homo sapien
25	91	3.5	1032	1 MT18_YEAST	P40469 saccharomyc
26	91	3.5	1054	1 POL_SIVMK	P05897 simlan immu
27	90	3.5	573	1 DPOL_MOUSE	Q9qxe2 mus musculu
28	90	3.5	948	1 RPOD_FODAN	O01521 podospora a
29	90	3.5	1125	1 TIE2_BOVIN	Q06807 bos taurus
30	90	3.5	1300	1 POL2_MOUSE	P11369 mus musculu
31	89.5	3.5	412	1 SY22_BACSU	P25151 bacillus su
32	89	3.5	678	1 UVRB_PASMU	P57844 pasteurella
33	88.5	3.4	286	1 PARB_CHLPN	Q9z7m0 chlamydia p

RESULT 1  
T3MH\_HAEIN  
ID T3MH\_HAEIN STANDARD; PRT; 629 AA.  
AC P71366;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Putative type III restriction-modification system HindVIP enzyme mod  
DE (EC 2.1.1.72) (HindVIP methyltransferase) (M.HindVIP).  
GN HI1056.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kervilavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RL Science 269:496-512(1995).  
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-  
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.  
CC -!- SUBUNIT: CONTAINS TWO DIFFERENT SUBUNITS: RES AND MOD. MOD IS  
CC A HOMOTETRAMER (BY SIMILARITY).  
CC -!- SIMILARITY: WITH OTHER TYPE III MOD PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U32786; AAC22721.1; -  
CC REBASE; 3701; M.Hindorf1056p.  
CC TIGR; HI1056; -  
CC InterPro; IPR002295; D2IN6\_mtfrase.  
CC InterPro; IPR001091; N4\_Mtase.  
CC InterPro; IPR002052; N6\_Mtase.  
CC InterPro; IPR002941; N6\_N4\_Mtase.  
CC Pfam; PF01555; N6\_N4\_Mtase; 1.  
CC PRINTS; PR00506; D2IN6MTFRASE.  
CC PRINTS; PR00508; S2IN4MTFRASE.  
CC PROSITE; PS00092; N6\_MTASE; 1.  
CC Hypothetical protein; Transferase; Methyltransferase; DNA-binding;  
KW

34 88.5 3.4 341 1 YF43\_MYCTU Q10783 mycobacteri  
35 88.5 3.4 536 1 DIT1\_YEAST P21623 saccharomyc  
36 88.5 3.4 859 1 ABR\_HUMAN Q12979 homo sapien  
37 88.5 3.4 949 1 RSCC\_ECOLI P14376 escherichia  
38 88 3.4 984 1 SECA\_AQUAE O67718 aquifex aeo  
39 88 3.4 1106 1 CIC2\_RABIT P13806 oryctolagus  
40 88 3.4 3137 1 CA36\_CHICK P15989 gallus gall  
41 87.5 3.4 476 1 ATPB\_GALSU Q08807 gaidieria s  
42 87.5 3.4 534 1 COX1\_SACDO P98001 saccharomyc  
43 87.5 3.4 557 1 COX1\_NEUCR P03945 neurospora  
44 87.5 3.4 577 1 SYMM\_CANAL O74634 candida alb  
45 87.5 3.4 627 1 MUTL\_BACSU P49850 bacillus su

ALIGNMENTS

KW Restriction system; Complete proteome.  
SQ SEQUENCE 629 AA; 71845 MW; 93ADAD90DDBA1E84 CRC64;

Query Match 3.9%; Score 101.5; DB 1; Length 629;  
Best Local Similarity 19.9%; Pred. No. 3.3;  
Matches 54; Conservative 50; Mismatches 98; Indels 69; Gaps 9;

QY 33 VNFNLSLQLDPSSIIHDGMDFLSKRLVLTARGSLPAFLFGKGRDTFLOFQNLNPAK 92  
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 198 IEQAYIKILMDDIEDFNFTNTVTVKTKIGGVSGSE-----GKSLKDSTEFINVFKNR 252  
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 93 SRGGPGDY-----YLKNYEDD-----IVRSVDVALDKQCKIAQHPDVMLE 134  
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 253 ERLFNPVQKTEVNEFINYEDGSKWKYTQVLIDLGEEKILLEKEGFKYHYHPN---- 308  
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 135 LQREKAQHMLVLLKESQNTSYNLIETPNPNN--YRTMHGRAVNGSOLGKDYIQLKSLLQ 192  
||| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 309 -----AQMTSIYKFSDQNLSKEIITYESHKYVRTNAOSSIRSKIIEDLYSIKNCIV 362  
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 193 PIRIYSRASLYGNIGRPKNVIALL-----DGFMKVAGSTV-----D 230  
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 363 SIE-----YIPQKGKAGNILLIEFYNASNKDMFMELSDLMIKEKKNYFVLQKVNTLWD 415  
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY 231 AVTWQHC-----YIDGRVVKVMDFLTRLLD 256  
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 416 DIQYNLNKGEYGIDFKNGKKPEALLRIID 446  
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 2

CYAA\_YEAST STANDARD; PRT; 2026 AA.

ID CYAA\_YEAST AC P08678;  
DC 01-JAN-1988 (Rel. 06, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenylyl  
cyclase)  
GN CYRI OR CDC35 OR HSR1 OR SRA4 OR YJL005W OR J1401.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OX Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
NCBI\_TaxID=4932;  
[1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=86079531; PubMed=2934138;  
RA Katakoka T., Broek D., Wigler M.;  
RT "DNA sequence and characterization of the S. cerevisiae gene encoding  
adenylate cyclase."  
RL Cell 43:493-505(1985).  
[2]  
RN SEQUENCE FROM N.A.  
RA To Van D., Perea J., Jacq C.;  
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
[3]  
RN SEQUENCE FROM N.A.  
RC STRAIN=S288C / FY1679;  
RA de Haan M., Smits P.H.M., Grivell L.A.;  
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
[4]  
RN SEQUENCE OF 1042-2026 FROM N.A.  
RX MEDLINE=88165073; PubMed=3327602;  
RA Masson P., Lenzen G., Jacquemin J.M., Danchin A.;  
RT "Yeast adenylate cyclase catalytic domain is carboxy terminal.";  
RL Curr. Genet. 10:343-352(1986).  
[5]  
RN MUTAGENESIS OF THR-1651.  
RX MEDLINE=91122042; PubMed=1991451;  
RA Pegar G., de Vendittis E., Vitelli A., Masturzo P., Zahn R.,  
RA Verrotti A.C., Kavounis C., Pal G.P., Fasano O.;  
RT "Identification of regulatory residues of the yeast adenylyl  
cyclase."  
RL EMBO J. 10:349-359(1991).



OS	symbiotic bacterium).	
OC	Bacteria; Proteobacteria; gamma subdivision; Buchnera.	
OX	NCBI_TaxID=1118099;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-TOKYO 1998;	
RX	MEDLINE=20445173; PubMed=10993077;	
RA	Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;	
RT	"Genome sequence of the endocellular bacterial symbiont of aphids	
RL	Buchnera sp. APS."	
CC	Nature 407:81-86(2000).	
CC	-1- CATALYTIC ACTIVITY: ATP + L-tryptophan + LRNA(Trp) = AMP +	
CC	diphosphate + L-tryptophanyl-trNA(Trp).	
CC	-1- SUBUNIT: HOMODIMER (BY SIMILARITY).	
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic.	
CC	-1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; AF001119; BAB13229.1; -	
DR	InterPro: IPR001412; trna-synt_L.	
DR	InterPro: IPR002306; trna-synt_Crp.	
DR	Pfam: PF00579; trna-synt_Lb; 1.	
DR	PRINTS; PR01039; TRNASYNTHTRP.	
DR	PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.	
KW	Aminoacyl-trna synthetase; protein biosynthesis; Ligase; ATP-binding;	
KW	Complete proteome.	
FT	SITE 14 22 "HIGH" REGION.	
FT	SITE 196 200 "RMSKS" REGION.	
FT	BINDING 199 199 ATP (BY SIMILARITY).	
FT	SEQUENCE 335 AA; 38549 MW; 871562D5A6734E3F CRC64;	
QY	Query Match 3.8%; Score 98.5; DB 1; Length 335;	
Db	Best Local Similarity 21.1%; Pred. No. 2.3; Mismatches 78; Indels 71; Gaps 10;	
Db	Matches 50; Conservative 38;	
QY	110 IVRSVDVALDKQCKIAHQDPVLMLEQREKAAQMHVLVLLKEQFSNTYSNLIETPNRYT 169	
Db	132 LMAADILLYQTNFVPGQDQKHVELTN-----IAHRENSLYGH-VFTLPKPLIT 181	
QY	170 MHGRAVNSQLGKDYIOLKSLQPIRYISRASLYGPNIGRPKNVIALLDG-----FMKVA 225	
Db	182 QHGS-----KIMSLPEPSKMSKSDI-----NKRNVIFLLDDIKTVISKIQ 222	
QY	226 GSTVDVAVTQHCYID-----GRVVKVM-----DFL-----KTRLDTL 258	
Db	223 NAYTDSPTPSKIYDIEKKPGISNLLBILSAITNKDIDILLKEGLMYSEFKNIVADHL 282	
QY	259 SDQIKTKQVNTVTPG-----KKTWLEGVTVTSAGGTNNISDSYAGFLWNLGML 311	
Db	283 SKFLYKLGKSYNDYRNDDEVLLKIAEGAMKSQLKSNKTLTKVY-----DKLGLI 332	
RESULT 5		
ID	CLPC_CHLPN STANDARD; PRT; 845 AA.	
AC	Q928A6; Q9K297; Q9JSE9;	
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Probable ATP-dependent CLP protease subunit.	
GN	CLP OR CPN0437 OR CP0316.	
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).	
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.	
OX	NCBI_TaxID=83558;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-CWL029;	
RX	MEDLINE=99206606; PubMed=10192388;	
RA	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,	
RT	Olinger L., Grimwood J., Davis R.W., Stephens R.S.; trachomatis."	
RL	Nat. Genet. 21:385-389(1999).	
CC	[2]	
CC	SEQUENCE FROM N.A.	
RC	STRAIN-AR39;	
RX	MEDLINE=20150255; PubMed=10684935;	
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., Bass S.,	
RT	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Dodson R.,	
RL	Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,	
CC	Ginn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,	
CC	Eisen J., Fraser C.M.;	
CC	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia	
CC	pneumoniae AR39.";	
CC	Nucleic Acids Res. 28:1397-1406(2000).	
CC	[3]	
CC	SEQUENCE FROM N.A.	
RC	STRAIN-J138;	
RX	MEDLINE=20330349; PubMed=10871362;	
RA	Shiral M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,	
RT	Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;	
RL	"Comparison of whole genome sequences of Chlamydia pneumoniae J138	
CC	from Japan and CWL029 from USA."	
CC	Nucleic Acids Res. 28:2311-2314(2000).	
CC	-1- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY. CLPC SUBFAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; AE001626; AAD18581.1; -	
DR	EMBL; AE002194; AAF38172.1; -	
DR	EMBL; AP002546; BAA98645.1; -	
DR	PHCI-2DPAGE; Q928A6; -	
DR	TIGR; CP0316; -	
DR	InterPro: IPR003593; AAA.	
DR	InterPro: IPR003959; AAA-subfam.	
DR	InterPro: IPR001270; CLP-AB.	
DR	InterPro: IPR004176; CLP-N.	
DR	InterPro: IPR001943; UVR.	
DR	Pfam: PF00004; AAA; 1.	
DR	Pfam: PF02861; CLP-N; 2.	
DR	Pfam: PF02151; UVR; 1.	
DR	PRINTS; PR00300; CLP-PROTEASEA.	
DR	SMART; SM00382; AAA; 2.	
DR	PROSITE; PS00870; CLPAB_1; 1.	
KW	Chaperone; ATP-binding; Repeat; Complete proteome.	
FT	NP_BIND 232 239 ATP (POTENTIAL).	
FT	NP_BIND 569 576 ATP (POTENTIAL).	
FT	CONFLICT 4 4 K -> T (IN REF. 3).	
FT	CONFLICT 141 141 K -> R (IN REF. 1).	
FT	SEQUENCE 845 AA; 94980 MW; C9D9205458812866 CRC64;	
QY	Query Match 3.8%; Score 97.5; DB 1; Length 845;	
Db	Best Local Similarity 19.7%; Pred. No. 10;	
Db	Matches 120; Conservative 85; Mismatches 228; Indels 175; Gaps 27;	
QY	8 VDRAAGLKEKTLILDVSTKNPV-----RTVNFNLSLQDPSIHDGWLDFLSKRLVTL 63	
Db	207 IGRSSEVERLILICRRKKNPNVLIGEAGVKTAVEGLAQKILNEVPDLRKRRLITL 266	
QY	64 ARGLSAPFLRFGG-----KRTDFLQF-----QNLRNP 90	
Db	267 DLALMIAGTKYRGQFEERIKAVMDEVKRGHGNILLFIDELHTIVGAGAAIDAISNLKP 326	









88	RNPAKSRGGPGDPYIYLKNYEDDVRSDVALQKGCKIAHQHPDMVLELQREKAAQ----	142
Qy	: :   :	
Ddb	: :   :	
84	RGKVLGKG-----FACKEY-----MTDLTNKKVYAANKIIHPSRVAKPHOREKDKEIELH	134
Qy	: :   :	
Ddb	: :   :	
143	-----MHLVLKKEQFSNTYSNLILTEPNRYTMHGRAVINGSQLGKDYIOLKSLLOP--IRI	196
Qy	: :   :	
Ddb	: :   :	
135	RILHHKHVOGYHYFEKENITYILEYCSRSM-----AHILK-----ARKVLTEPEVRY	184
Qy	: :   :	
Ddb	: :   :	
197	YSRASLYXGNIGRPKNRIALLDGFMKVGAGSTVDAVTWQHCYIDGRVVKMVDFLKTRLLD	256
Qy	: :   :	
Ddb	: :   :	
185	YLROIVSGLKYLHEQE-----ILHRDLKLGNFFIN-----EAMELKVKGDGFLGAARLE	231
Qy	: :   :	
Ddb	: :   :	
257	TLSDOIIRKIQKVNVTYP-----GKKTLWEGVV--TTSAG-----GTNNLSDSY--	298
Qy	: :   :	
Ddb	: :   :	
232	PLEHRRRTTCGTPNYLSPEVLNKQGHCESDIWALGCVMYTMLLRGPFPFTTNLKETYRC	291
Qy	: :   :	
Ddb	: :   :	
299	--AAGFLWLNTL-----GMLAQ-----GIDVTRHSFPDHGY--NHLVDQNFNPL	340
Qy	: :   :	
Ddb	: :   :	
292	IREARYTMPSSLLAPAKHLIASMLSKNPBDRPSLDIIDIRHDFYFLOGFTPORLSSSCCHTV	351
Qy	: :   :	
Ddb	: :   :	
341	PBYWTSLATVXPRIGPKVIHVHVAGLORKPR-----PGRVIRDKLRIYAHCTNHNNHNHYVRG	396
Qy	: :   :	
Ddb	: :   :	

[illegible]



use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L03524; AAA28151.1; .  
EMBL; Z78412; CAB01648.1; ALT\_INIT.  
HSP; P11362; IFGK.  
WormPep; M176.6; KCL2470.  
InterPro; IPR00719; Euk\_pkinase.  
InterPro; IPR001245; Tyr\_pkinase.  
Pfam; PF00069; pkinase; 2.  
SMART; SM00219; TyrcK; 1.  
PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
Transferase; Tyrosine-protein kinase; ATP-binding; Receptor;  
Transmembrane; Glycoprotein; Signal.  
SIGNAL 1 26 POTENTIAL.  
CHAIN 27 488 RECEPTOR-LIKE TYROSINE-PROTEIN KINASE  
KIN-15.  
DOMAIN 27 50 EXTRACELLULAR (POTENTIAL).  
TRANSMEM 51 70 POTENTIAL.  
DOMAIN 71 488 CYTOPLASMIC (POTENTIAL).  
DOMAIN 144 458 PROTEIN KINASE.  
NP\_BIND 150 158 ATP (BY SIMILARITY).  
BINDING 183 183 ATP (BY SIMILARITY).  
ACT\_SITE 319 319 BY SIMILARITY.  
CARBOHYD 25 25 N-LINKED (GLCNAC. .) (POTENTIAL).  
SEQUENCE 488 AA; 56544 MW; F891B9BEEA128977 CRC64;

Query Match 3.6%; Score 93; DB 1; Length 488;  
Best Local Similarity 21.2%; Pred. No. 10;  
Matches 82; Conservative 49; Mismatches 141; Indels 114; Gaps 21;

QY 21 LLDVSTKN--PV--RTVNENFLSLQDPSI-IHDGWL-----DFLSSRLVTLARGLSPA 70  
Db 106 LLDVTSVNEETIVKRPINERITENLEFDPFIDQAKLEISDKLGSFFGEVCYGL--L 163  
QY 71 FLRFQKRTDFLOFQNLNPAKSGGPGDYLLKYNVEDDIVRSVDVALDKQKCKTAQHPD 130  
Db 164 SMRTSNTETDTL--QKLSVAVQSNPDPTQENOEKMIED-----ETKLMCAITGRNP 212  
QY 131 VM-----LELQREKAQMLV-----LLK--EFSNTYSNLIILTEPNNYRTMHGRAVN 176  
Db 213 ILAIGAVTANGSARNLLIVEFEGCDLLKFLKESKIFKDELVEKNGY-----263  
QY 177 GSOLGRDYIQLKSLLOPIRIYSRASLYGPNIGRPKNVI-ALLDGFPMKVGSTVDVATWQ 235  
Db 264 -----LLPKSIRKTYMFEN-----EDDVIEESLDSL-----CTSLLSFS 300  
QY 236 HCVIDG-----RVKVMDFLKLRLDLSQIRKIKQVNT 271  
Db 301 YQIAEGMEYLASTPCVHRDLALRNVLNKNKTIRIADFLGRKQYV--DGYRTIKGVGT 358  
QY 272 YTPGKKIWLGVVYTSAGGTNNLSD--SYAAGFLMNTLGLMLANOGI--DVVIRHSFED 326  
Db 359 PMPAR--WMAPEVWRE--GKCTESDWSYGVSYLMEFSLGELPYSVNSDV-----407  
QY 327 HGNHVLVDQNFNPLPDYWLSLLYKRL 352  
Db 408 --FEHVVGQNLPMPOYCHPKMYDRM 431

RESULT 14  
ID PHSG\_SYNY3 STANDARD; PRT; 849 AA.  
AC P73511;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Glycogen phosphorylase (EC 2.4.1.1).

GN GLGP OR SLL1356.  
OS Synchocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97061201; PubMed=8905231;  
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hiroseawa M., Sugliura M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
RA Shimpō S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
RA Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
CC -!- FUNCTION: PHOSPHORYLASE IS AN IMPORTANT ALLOSTERIC ENZYME IN  
CC CARBOHYDRATE METABOLISM. ENZYMES FROM DIFFERENT SOURCES DIFFER IN  
CC THEIR REGULATORY MECHANISMS AND IN THEIR NATURAL SUBSTRATES.  
CC HOWEVER, ALL KNOWN PHOSPHORYLASES SHARE CATALYTIC AND STRUCTURAL  
CC PROPERTIES (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: {(1.4)-alpha-D-glucosyl}(N) + phosphate =  
CC {(1.4)-alpha-D-glucosyl}(N-1) + alpha-D-glucose 1-phosphate.  
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE GLYCOGEN PHOSPHORYLASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; D90907; BAAL7551.1; .  
HSP; P06738; IYGP.  
InterPro; IPR000811; Phosphorylase.  
Pfam; PF00343; Phosphorylase; 1.  
PROSITE; PS00102; PHOSPHORYLASE; 1.  
Transferase; Glycosyltransferase; Carbohydrate metabolism;  
KW Glycogen metabolism; Pyridoxal phosphate; Complete proteome.  
FT BINDING 679 679 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
FT SEQUENCE 849 AA; 96648 MW; 657D2C5B83C23A3D CRC64;

Query Match 3.6%; Score 93; DB 1; Length 849;  
Best Local Similarity 19.8%; Pred. No. 22;  
Matches 94; Conservative 51; Mismatches 130; Indels 200; Gaps 25;

QY 32 TVNENFLSLQDPSIHDGWLDFLSSRLVTLARGLSPAFLRFQKRTDFLOFQNLNPA 91  
Db 465 TILKDFYEL-----WPEKFSNK-----TNGVTP-----RRWVLNPNRLSNLI 502  
QY 92 KSRGPGPDYVLKNVEDDIVR-----SDVALDKQKCKTAQ-----HP 129  
Db 503 SSRIGDG---WIKNL-DELKQLEPADLAGFRQDWCKVKREVQKDLARYIHTRTDLVVP 558  
QY 130 DYMLELQ-----REKAAQHMLVLLKEQFSNTYSNLIILTEPNNYRTMHGRAVNSOLG 181  
Db 559 DSLFDVQVKRIHEYKRQHLNLIHLVHLVLTQKNN--PNLDVT--PRTF-----IYGGKAA 609  
QY 182 KDYIQLKSLLOPIRIYSRASLYGPNIGRPKNVIALLDGFPMKVGSTVDVATWQHICYIDG 241  
Db 610 PGYFTAKLIIKLINSVADVNNNDPTIG--DLKVIPLDYNVK-----FGQ 653  
QY 242 RVVKVMDFLKRLDLSQIRKIKQVNTYTPGKKIMLEGVVTTISAGTNNLSDSYAAG 301  
Db 654 RVYPAAD-----LSEQIS-----TAGKE-----ASGTGNMKFMSMGA 685  
QY 302 FLWLNTLGLANQIGIDV-----318  
Db 686 L-----TIGTLDGANIEIREEVGAENFFLEGLTTPVEKTLASGYQPYEYNNNNANKAVV 741

```
QY 319 -VIRHSFEDGHNHLDQNFNPLDPYWLWLLYKRLIGPKVLAVHVGLOKPRGVRVD 377
Db 742 DLINGGFFSHGDTAL-----FRPLMD-----SLGQD-----PYLVFAD 775
QY 378 KLRIYAHCTNH-----HHNHVVRGSIWLFINLHRSKKIKLACTLRDKLKHQY 426
Db 776 -FOAYVDCNQVGEAYKDOENWAR-----MAINVARWGK-----FSSDRTIREY 819

RESULT 15
SNK_MOUSE STANDARD; PRT; 682 AA.
AC PS3351;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Serine/threonine-protein kinase SNK (EC 2.7.1.-) (Serum inducible
DE kinase).
DE SNK.
GN SNK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92375085; PubMed=1508211;
RA Simmons D.L., Neel B.G., Stevens R., Evett G., Erikson R.L.;
RT "Identification of an early-growth-response gene encoding a novel
RT putative protein kinase.";
RL Mol. Cell. Biol. 12:4164-4169(1992).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE DIVISION OF AT LEAST SOME CELL
CC TYPES, SUCH AS FIBROBLASTS, AND COULD FUNCTION IN EMBRYOGENESIS,
CC WOUND HEALING OR NEOPLASIA.
CC -!- TISSUE SPECIFICITY: BRAIN, LUNG AND HEART.
CC -!- INDUCTION: BY SERUM AND PHORBOL ESTER.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CDC5/POLO SUBFAMILY.
CC -!- SIMILARITY: CONTAINS 2 POLO BOX DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. Usage by and for commercial
CC or send an email to license@sib-sib.ch.
CC -----
DR EMBL; M96163; -; NOT_ANNOTATED_CDS.
DR HSPSP; P00518; 1PHK.
DR MGD; MGI:1099790; Snk.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000959; POLO_box.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00659; POLO_box; 2.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS50078; POLO_BOX; 2.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat.
FT DOMAIN 54 59 POLY-HIS.
FT DOMAIN 79 331 PROTEIN KINASE.
FT NP_BIND 85 93 ATP (BY SIMILARITY).
FT BINDING 108 108 ATP (BY SIMILARITY).
FT ACT_SITE 202 202 BY SIMILARITY.
FT DOMAIN 507 570 POLO BOX 1.
FT DOMAIN 603 674 POLO BOX 2.
SQ SEQUENCE 682 AA; 77811 MW; 586DEABFD7208A9D CRC64;
```

Query Match 3.68; Score 92.5; DB 1; Length 682;  
Best Local Similarity 20.9%; Pred. No. 18;

```
Matches 89; Conservative 58; Mismatches 167; Indels 111; Gaps 19;
QY 88 RNPASRGSGPDYLYLKNYEDDIVRSDVALDKQKGIQAHPDVMLELQREKAAQ-----142
Db 81 RGVKLGKG-----FAKCYE-----MTDLTNKVVAAKIIPHRSVAKPHOREKIDKEIELH 131
QY 143 -----MHLVLLKEQFSNTYSNLITPEPNRYRTMHGRAVNGSOLGKDYIOLKSLQP-IRI 196
Db 132 RLLHHKHVVQFYHFEDEKENIYILLEYCSRSM-----AHILK-----ARKVLTEPEVRY 181
QY 197 YSRASIYGNIGRPRKNVTALLDGFPMKVAGSTVDATWQHCVYIDGRVVKVMDFLKTRILD 256
Db 182 YLRQIVSGLKYLHEQE-----ILHRLKLGNNFIN-----EAMELVGDFGLAARLE 228
QY 257 TLSDQIRKIQKVNTYTP-----GKKIWLEGV--TTSAG-----GTNNLSDSY-- 298
Db 229 PLEHRRRTICGTPNVLSPVLNKGHCESDIWALGCVMYTMLLGGPPPTTNLRETYRC 288
QY 299 ---AAGFLMLNTL-----GMLANQ-----GIDVVIHRSFFDHGY--NHLVDQNFNPL 340
Db 289 IREARYTMPSSLLAPAKHLLIASMLSKNPDPSLDDIIRHDFFLQGFDPDLSSSCCHTV 348
QY 341 PDYWLSSLYKRLIGPKVLAVHVGLOKPRGVRVDKLRIVAHCTNHHNHVVRGSIYL 400
Db 349 PDFHLSSPAKNFF-KKAAAALFEGGKKDKAR-----YNDTHNKVSK 387
QY 401 FTINLHRSRKIKLAGTLRDKLVHVOYLLQPY-----GOEGLKSKSVQLNGOPLVMV 451
Db 388 EDEDIYKLRHDLKKVSIQOQSKHRADEEPQPPPTTVARSGETSAVENK--QQIGDAIRMI 445
QY 452 DDGTL 456
Db 446 VRGTL 450

Search completed: July 30, 2002, 08:31:30
Job time: 977 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:30:43 ; Search time 123.94 Seconds  
(without alignments)  
688.127 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_42\_534  
Perfect score: 2577  
Sequence: 1 GDRRLPVDRAAGLKEKTLI.....PVTMGFFVKNVNALACRYR 493

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_19:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	2577	100.0	534	4	Q9HB38	Q9hb38 homo sapien
2	2538	98.5	592	4	Q9HB37	Q9hb37 homo sapien
3	2270	88.1	480	4	Q9HB39	Q9hb39 homo sapien
4	953	37.0	536	11	Q9QZF8	Q9qzf8 rattus norv
5	927.5	36.0	545	4	Q9UL39	Q9ul39 homo sapien
6	926.5	36.0	543	4	Q9Y251	Q9y251 homo sapien
7	915.5	35.5	545	6	Q9MYV0	Q9myv0 bos taurus
8	835.5	32.4	523	13	Q90YK5	Q90yk5 gallus gall
9	302	11.7	521	10	Q9SDA1	Q9sdai arabidopsis
10	302	11.7	543	10	Q9FF10	Q9ff10 arabidopsis
11	261.5	10.1	516	10	Q9FLK8	Q9flk8 arabidopsis
12	256.5	10.0	527	10	Q9LRC8	Q9lrc8 scutellaria
13	246	9.5	536	10	Q9FZP1	Q9fzp1 arabidopsis
14	171	6.6	935	5	Q9VE79	Q9ve79 drosophila
15	155	6.0	190	10	Q82604	Q82604 arabidopsis
16	113	4.4	4533	5	Q9BIX3	Q9bix3 tetrahymena

17	111	4.3	1260	5	015639	015639 dictyosteli
18	110	4.1	174	10	Q9ATW5	Q9atw5 zea mays (m
19	106.5	4.1	885	5	061232	061232 lymnaea sta
20	106	4.1	768	3	Q9P8N9	Q9p8n9 cladosporiu
21	105	4.1	1020	5	025415	025415 leishmania
22	101.5	3.9	1234	5	Q9Y1H7	Q9y1h7 dictyosteli
23	100.5	3.9	582	17	Q9HRZ5	Q9hrz5 halobacteri
24	99.5	3.9	455	3	Q12229	Q12229 saccharomyc
25	99.5	3.9	518	5	Q968Y8	Q968y8 caenorhabdi
26	99.5	3.9	535	5	Q968Y7	Q968y7 caenorhabdi
27	99.5	3.9	1272	16	Q98PR8	Q98pr8 mycoplasma
28	99	3.8	1713	9	Q94MA1	Q94mal lactococcus
29	99	3.8	2403	5	015881	015881 trypanosoma
30	99	3.8	4578	13	042181	042181 fugu rubrip
31	98.5	3.8	1214	12	Q9INJ1	Q9inj1 banna virus
32	98	3.8	491	2	008457	008457 clostridium
33	98	3.8	907	10	Q9SE90	Q9se90 brassica ol
34	98	3.8	1090	12	Q91E95	Q91e95 human rotav
35	98	3.8	1281	11	088915	088915 mus musculu
36	97.5	3.8	515	5	Q24109	Q24109 drosophila
37	97.5	3.8	553	5	Q9VEQ5	Q9veq5 drosophila
38	97.5	3.8	1167	16	Q9A205	Q9a205 streptococc
39	97	3.8	838	10	Q9FG37	Q9fg37 arabidopsis
40	96.5	3.7	1429	2	Q9LAP7	Q9lap7 alteromonas
41	96	3.7	763	10	Q9ZUL5	Q9zul5 arabidopsis
42	96	3.7	898	10	Q9SIB9	Q9sib9 arabidopsis
43	96	3.7	1219	11	Q9QYM3	Q9qym3 mus musculu
44	95.5	3.7	406	2	Q57513	Q57513 coxiella bu
45	95.5	3.7	472	16	Q98QF0	Q98qf0 mycoplasma

ALIGNMENTS

RESULT 1

Q9HB38	PRELIMINARY;	PRT;	534 AA.
ID	Q9HB38		
AC	Q9HB38;		
DT	01-MAR-2001 (TremBLrel. 16, Created)		
DT	01-MAR-2001 (TremBLrel. 16, Last sequence update)		
DT	01-MAR-2001 (TremBLrel. 16, Last annotation update)		
DE	HEPARANASE-LIKE PROTEIN HPA2B.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20483645; PubMed=11027606;		
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,		
RA	Hircock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;		
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian		
RT	Heparanase Family Member.;"		
RL	Biochem. Biophys. Res. Commun. 276:1170-1177(2000).		
SQ	EMBL; AF282886; AAG23422.1; -.		
DR	SEQUENCE 534 AA; 60663 MW; C3DE5E900CB338C4 CRC64;		

Query Match 100.0%; Score 2577; DB 4; Length 534;  
Best Local Similarity 100.0%; Pred. No. 1.2e-198;  
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNVNENFLSLQLDPSIIHGWLDLSSKRL	60
Db	42	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNVNENFLSLQLDPSIIHGWLDLSSKRL	101
QY	61	VTLAGRLSPAFRLFGKRTDFLQFNLRNPAKSRGPGDYLLKNYEDDIVRSDVALDKQ	120
Db	102	VTLAGRLSPAFRLFGKRTDFLQFNLRNPAKSRGPGDYLLKNYEDDIVRSDVALDKQ	161
QY	121	KCKTAQHPDVMLEQLQREKAAQMHVLVLLKEQFSNTYSNLILTEPNRYTMHGRAVNGSOL	180
Db	162	KCKTAQHPDVMLEQLQREKAAQMHVLVLLKEQFSNTYSNLILTEPNRYTMHGRAVNGSOL	221

Qy	181	GKDYIQLKSLLOPIRIYSRASLYGPNIGRPRKNNVIALDGFMKVAGSTVDVATWQHCHYID	240
Db	222	GKDYIQLKSLLOPIRIYSRASLYGPNIGRPRKNNVIALDGFMKVAGSTVDVATWQHCHYID	281
Qy	241	GRVVKVMDFLKRLDLSDOIIRKIQKVVNTYTPGKKIWLGVVTTSGAGTNNLSDSYAA	300
Db	282	GRVVKVMDFLKRLDLSDOIIRKIQKVVNTYTPGKKIWLGVVTTSGAGTNNLSDSYAA	341
Qy	301	GFLWLTNLGLANOGIDVVIIRHSFFDHGYNHLVDQNFNPLDPYWLSSLYKRLIGPKVLAV	360
Db	342	GFLWLTNLGLANOGIDVVIIRHSFFDHGYNHLVDQNFNPLDPYWLSSLYKRLIGPKVLAV	401
Qy	361	HVAGLQRPGRVIRDKLRIYAHCTNNHNNHNVYRGSITLFIINLHRSRKKIKLAGTLRD	420
Db	402	HVAGLQRPGRVIRDKLRIYAHCTNNHNNHNVYRGSITLFIINLHRSRKKIKLAGTLRD	461
Qy	421	KLVHQLLPYQOGELKSKSVOLNGQPLVMVDDGTLPELKPRPLRAGRTLVIPTVMGFF	480
Db	462	KLVHQLLPYQOGELKSKSVOLNGQPLVMVDDGTLPELKPRPLRAGRTLVIPTVMGFF	521
Qy	481	VYKVNALACRYR 493	
Db	522	VYKVNALACRYR 534	
RESULT	2		
Q9HB37			
ID	Q9HB37	PRELIMINARY;	PRT; 592 AA.
AC	Q9HB37;		
DT	01-MAR-2001 (Tremblrel. 16, Created)		
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)		
DT	01-MAR-2001 (Tremblrel. 16, Last annotation update)		
DE	HEPARANASE-LIKE PROTEIN HPA2C.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20483645; PubMed=11027606;		
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,		
RA	Hircock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;		
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian		
RT	Heparanase Family Member."		
RL	Biochem. Biophys. Res. Commun. 276:1170-1177(2000).		
DR	EMBL; AF282887; AAG23423.1; -.		
SQ	SEQUENCE 592 AA; 66580 MW; 95C384AD9A74258E CRC64;		
Query Match	98.5%;	Score 2538;	DB 4; Length 592;
Best Local Similarity	89.5%;	Pred. No. 1.9e-195;	
Matches 493;	Conservative 0;	Mismatches 0;	Indels 58; Gaps 1;
Qy	1	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFSLQLDPSIIHDGWLDFLSSKRL	60
Db	42	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFSLQLDPSIIHDGWLDFLSSKRL	101
Qy	61	VTIARGLSPAFIRFGGKRTDFLQFNLRNPAKSRGGPGDPYILKNYEDDIVRSDVALDKQ	120
Db	102	VTIARGLSPAFIRFGGKRTDFLQFNLRNPAKSRGGPGDPYILKNYEDDIVRSDVALDKQ	161
Qy	121	KGCKIAQHPDVMLEQREKAQHLLVLLKEQFSNTYSNLITL-----	162
Db	162	KGCKIAQHPDVMLEQREKAQHLLVLLKEQFSNTYSNLITL-----	221
Qy	163	-----EPNNYRTMHGRAVNGSOLGK	192
Db	222	IFALNALRRPNNSWSSSALSLLKYSASKYNIWSWELGNEPNPNYRTMHGRAVNGSOLGK	281
Qy	183	DYIQLKSLLOPIRIYSRASLYGPNIGRPRKNNVIALDGFMKVAGSTVDVATWQHCHYIDGR	242
Db	282	DYIQLKSLLOPIRIYSRASLYGPNIGRPRKNNVIALDGFMKVAGSTVDVATWQHCHYIDGR	341

Qy	243	VVKVMDFLKRLDLSDOIIRKIQKVVNTYTPGKKIWLGVVTTSGAGTNNLSDSYAAGF	302
Db	342	VVKVMDFLKRLDLSDOIIRKIQKVVNTYTPGKKIWLGVVTTSGAGTNNLSDSYAAGF	401
Qy	303	LWLTNLGLANOGIDVVIIRHSFFDHGYNHLVDQNFNPLDPYWLSSLYKRLIGPKVLAVH	362
Db	402	LWLTNLGLANOGIDVVIIRHSFFDHGYNHLVDQNFNPLDPYWLSSLYKRLIGPKVLAVH	461
Qy	363	AGLQRPGRVIRDKLRIYAHCTNNHNNHNVYRGSITLFIINLHRSRKKIKLAGTLRDKL	422
Db	462	AGLQRPGRVIRDKLRIYAHCTNNHNNHNVYRGSITLFIINLHRSRKKIKLAGTLRDKL	521
Qy	423	VHQLLPYQOGELKSKSVOLNGQPLVMVDDGTLPELKPRPLRAGRTLVIPTVMGFFV	482
Db	522	VHQLLPYQOGELKSKSVOLNGQPLVMVDDGTLPELKPRPLRAGRTLVIPTVMGFFV	581
Qy	483	KVNALACRYR 493	
Db	582	KVNALACRYR 592	
RESULT	3		
Q9HB39			
ID	Q9HB39	PRELIMINARY;	PRT; 480 AA.
AC	Q9HB39;		
DT	01-MAR-2001 (Tremblrel. 16, Created)		
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)		
DT	01-MAR-2001 (Tremblrel. 16, Last annotation update)		
DE	HEPARANASE-LIKE PROTEIN HPA2A.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20483645; PubMed=11027606;		
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,		
RA	Hircock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;		
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian		
RT	Heparanase Family Member."		
RL	Biochem. Biophys. Res. Commun. 276:1170-1177(2000).		
DR	EMBL; AF282887; AAG23421.1; -.		
SQ	SEQUENCE 480 AA; 53900 MW; F75F89F67AC1FF83 CRC64;		
Query Match	88.1%;	Score 2270;	DB 4; Length 480;
Best Local Similarity	89.0%;	Pred. No. 4.8e-174;	
Matches 439;	Conservative 0;	Mismatches 0;	Indels 54; Gaps 1;
Qy	1	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFSLQLDPSIIHDGWLDFLSSKRL	60
Db	42	GDRRLPVDRAAGLKEKTLILLDVSTKNPVRTVNENFSLQLDPSIIHDGWLDFLSSKRL	101
Qy	61	VTIARGLSPAFIRFGGKRTDFLQFNLRNPAKSRGGPGDPYILKNYEDDIVRSDVALDKQ	120
Db	102	VTIARGLSPAFIRFGGKRTDFLQFNLRNPAKSRGGPGDPYILKNYEDDIVRSDVALDKQ	149
Qy	121	KGCKIAQHPDVMLEQREKAQHLLVLLKEQFSNTYSNLITLPEPNYRTMHGRAVNGSOL	180
Db	150	-----EPNNYRTMHGRAVNGSOL	167
Qy	181	GKDYIQLKSLLOPIRIYSRASLYGPNIGRPRKNNVIALDGFMKVAGSTVDVATWQHCHYID	240
Db	168	GKDYIQLKSLLOPIRIYSRASLYGPNIGRPRKNNVIALDGFMKVAGSTVDVATWQHCHYID	227
Qy	241	GRVVKVMDFLKRLDLSDOIIRKIQKVVNTYTPGKKIWLGVVTTSGAGTNNLSDSYAA	300
Db	228	GRVVKVMDFLKRLDLSDOIIRKIQKVVNTYTPGKKIWLGVVTTSGAGTNNLSDSYAA	287
Qy	301	GFLWLTNLGLANOGIDVVIIRHSFFDHGYNHLVDQNFNPLDPYWLSSLYKRLIGPKVLAV	360
Db	288	GFLWLTNLGLANOGIDVVIIRHSFFDHGYNHLVDQNFNPLDPYWLSSLYKRLIGPKVLAV	347



QY 361 HVAGLQKPRPRGRVTRDKLRIYAHCTNNHHNNHNVYRGSTLFIINLHRSRKKIKLAGTLRD 420  
|||||  
Db 348 HVAGLQKPRPRGRVTRDKLRIYAHCTNNHHNNHNVYRGSTLFIINLHRSRKKIKLAGTLRD 407  
|||||  
QY 421 KLVHQLLPQYGOEGLSKSVOLNGOPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFF 480  
|||||  
Db 408 KLVHQLLPQYGOEGLSKSVOLNGOPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFF 467  
|||||  
QY 481 VVKYNALACRYR 493  
|||||  
Db 468 VVKYNALACRYR 480  
|||||  
RESULT 4  
Q9QZF8 PRELIMINARY; PRT; 536 AA.  
AC Q9QZF8;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE HEPARANASE.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Podyma K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;  
RT "Heparanase from parathyroid cell line.";  
RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF184967; AAF04563.1;  
SQ SEQUENCE 536 AA; 60569 MW; 6208B1FD9EE28421 CRC64;

Query Match 37.0%; Score 953; DB 11; Length 536;  
Best Local Similarity 39.3%; Pred. No. 3.9e-68;  
Matches 216; Conservative 83; Mismatches 165; Indels 86; Gaps 10;  
QY 1 GDRRLPVDRAAG-LKEKTLILLDVSTKNPVRTNENFLSLDPSIHD-GWLDPLSSK 58  
|||  
Db 12 GLRLALQTGTAGTAPTKDQVLEFYTKRLFQSVSPSLTIDASLATDPRFLFLGSP 71  
|||  
QY 59 RLVTIARGSLPFLRFGKRTDQLQNLNPAKSRGGPGPDYLYKNYEDDILVRSVDALD 118  
|||  
Db 72 RLRLARGSLPFLRFGKRTDQLQNLNPAKSRGGPGPDYLYKNYEDDILVRSVDALD 123  
|||  
QY 119 KQKCKIAQHPDYMELQREKAAQHLVLLKEQF-----SNTYS----- 157  
|||  
Db 124 -----RVSADVLRLKQWEPFQ-ELLLRLQYQREFKNSTYSRSDVDMLYSFAKCSRL 175  
|||  
QY 158 -----NLIL-----TEPNRYTMHGRAVNGSOL 180  
|||  
Db 176 DLIFGLNALLRTPDLRWSSNAQLLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQL 235  
|||  
QY 181 GKDYIOLKSLPIRIYSRASLYGNIPRKNVIALDGFPMKVGSTVDVATWQHCHYD 240  
|||  
Db 236 GEDFVELHKLQK-SAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVDSLTWHHYLN 294  
|||  
QY 241 GRVVKVMDFLKRLDILDSQIRKIQVYNTYTPGKKIWLGVVTTTSAGTNNLSDSYAA 300  
|||  
Db 295 GRVATKEDFLSSDVLDTFLSQKILKVTKEMTGPKKWLGETSSAYGGGAPLLSNTFAA 354  
|||  
QY 301 GFLWLNTLGMLANQIDVIRHSFFDHGYNHLVDQNFNPLPDYWSLLYKRLIGPKVLAV 360  
|||  
Db 355 GFWMLDKGLSQAQIEVVRQVFGAGNYHLVDENFPLPDYWSLLYKRLIGPKVLAV 414  
|||  
QY 361 HVAGLQKPRPRGRVTRDKLRIYAHCTNNHHNNHNVYRGSTLFIINLHRSRKKIKLAGTLRD 420  
|||||  
Db 415 RVKQPD-----RSKLRVYLHCTNVYHPRYREGDILYVLNLHNVTKHLKLPMPMS 465  
|||||  
QY 421 KLVHQLLPQYGOEGLSKSVOLNGOPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFF 480  
|||||

Db 466 RVDYLLKPFSGDGLLSKSVOLNGQTLKMDVDEQTLPALTEKPLPAGSSLSYPAFSYGF 525  
|||||  
QY 481 VVKYNALAC 490  
|||  
Db 526 VIRNAKIAAC 535  
|||  
RESULT 5  
Q9UL39 PRELIMINARY; PRT; 545 AA.  
AC Q9UL39;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE HEPARANASE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RX MEDLINE=20229546; PubMed=10764835;  
RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;  
RT "Heparanase expression in invasive trophoblasts and acute vascular damage.";  
RL Glycobiology 10:467-475(2000).  
DR EMBL; AF084467; AAD54516.1;  
SQ SEQUENCE 545 AA; 61418 MW; 67B80ACD73C5A9A1 CRC64;  
Query Match 36.0%; Score 927.5; DB 4; Length 545;  
Best Local Similarity 38.9%; Pred. No. 4.5e-66;  
Matches 208; Conservative 78; Mismatches 162; Indels 87; Gaps 10;  
QY 16 EKTLLLDVSTKNPVRTNENFLSLDPSIHD-GWLDPLSSKRLVTLARGSLPFLR 74  
|||  
Db 37 QQDVLDLDEFTQERPLHLVSPSFLSVTIDANLATDPRFLILGSPKRLTARGSLPFLR 96  
|||  
QY 75 GSKRTDQLQNLNPAKSRGGPGPDYLYKNYEDDILVRSVDALDQKCKIAQ-HPDVML 133  
|||  
Db 97 GGTGTDLFIF----DPKKESTFEERSYQSQVNQDI-----CKYGIIPDVEE 140  
|||  
QY 134 ELQREKAAQHLVLLKEQF-----SNTYS-----NLIL----- 161  
|||  
Db 141 KRLLEWPIQEQ-LLEHYQKKFNSTYSRSSVDVLYTFANCSGLDIFGLNALLRTADL 199  
|||  
QY 162 -----TEPNRYTMHGRAVNGSOLGKDYIOLKSLQPIR 195  
|||  
Db 200 QWNSSNAQLLLDYCSSKGYNISWELGNEPNSFLKADIFINGSOLGEDFQLHKLRLK-S 258  
|||  
QY 196 IYSRASLYGNIPRKNVIALDGFPMKVGSTVDVATWQHCHYDGRVVKVMDFLKTRL 255  
|||  
Db 259 TFKNAKLYGPDIGQPRGKTKAKMLKSLKAGGEVIDSVTHHYLYNGRTATREDFLAPDVL 318  
|||  
QY 256 DTLSQDQIRKIQVYNTYTPGKKIWLGVVTTTSAGTNNLSDSYAAAGFLMLNTLGMLANQ 315  
|||  
Db 319 DIFISSQVQVQVVESTRPGKKVWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARWG 378  
|||  
QY 316 IDVIRHSFFDHGYNHLVDQNFNPLPDYWSLLYKRLIGPKVLAVHAGLQKPRGRVI 375  
|||  
Db 379 IEVVMRQVFGAGNYHLVDENFPLPDYWSLLYKRLIGPKVLAVHAGLQKPRGRVI 431  
|||  
QY 376 RDKLRIYAHCTNNHHNNHNVYRGSTLFIINLHRSRKKIKLAGTLRDLKLVHQLLPQYGOE 435  
|||  
Db 432 --KLRVYLHCTNDNPRYKEGDILYVLNLHNVTKHLKLPMPMS 489  
|||  
QY 436 LKSKSVOLNGOPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFFVVKYNALAC 490  
|||||  
Db 490 LLSKSVOLNGOPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFFVVKYNALAC 544  
|||||

[illegible]

```
Db 311 DFLNPDLIDTFISSVQKTLRIVEKIRPLKKVWLGWLTSSAFGGGAPFLSNTPFAAGFWMLDK 370
Qy 308 LGLMANQIDVYIRHSFHDGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVAGLQR 367
Db 371 LGLSARMGIEVVMQVLEFAGNYHLVDGNFEPDPYWLSSLLFKLVGNKVLWASVKGPD- 429
Qy 368 KPRGRVTRDKLRIYAHCTNHHNNHNYVSGSITLFIINLHRSRKKIKLAGTLRDKLVHQL 427
Db 430 -----RSFRVYLHCTNTHKPRYKEDGTLFALNHLNHTKHELPHLFNRKQVDKYL 481
Qy 428 LPQYGOGLKSKSVOLGNQPLVMVDGTLPELKPRLRAGRTLVIPTVMTGFFVYKVNNA 487
Db 482 IKPSTGDLKSKSVOLGNQILKXWDEQTLPALTEPLHPGSSLGMPFPYSYGFVIRNAKV 541
Qy 488 LAC 490
Db 542 AAC 544

RESULT 8
ID Q90YK5 PRELIMINARY; PRT; 523 AA.
AC Q90YK5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HEPARANASE.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11387326;
RA Goldsmith O., Zcharia E., Aingorn H., Guatta-Rangini Z., Atzmon R.,
RA Michal I., Pecker I., Mitrani E., Vlodaysky I.,
RT "Expression Pattern and Secretion of Human and Chicken Heparanase Are
RT Determined by Their Signal Peptide Sequence.";
RL J. Biol. Chem. 276:29178-29187(2001).
DR EMBL; AY037007; AAK82648.1; -.
SQ SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF881 CRC64;

Query Match 32.4%; Score 835.5; DB 13; Length 523;
Best Local Similarity 36.6%; Pred. No. 1.1e-58;
Matches 192; Conservative 74; Mismatches 180; Indels 79; Gaps 8;

Qy 22 LDVSTKNPVRTVNEFLSLQDPSIIHD-GWLDFLSSKRLVTLARGLSPAFLPFGKRTD 80
Db 21 LQGLREPIGAVSPAFSLTLTDLASLARDPRFVALLRHPKLHTLASGLSPGLFRFGTSTD 80
Qy 81 FLQFNLRNPAKSRGGPGPDYLYKNVEDDIVRSDVALD-----KQ 120
Db 81 FLIF-----NPNKD-----STWEKVLSEFOAKDCEAMPFVVPKLLLTQWPIQ 126
Qy 121 KCKCIAOHP-----DVM-----LELREKAQMHL 145
Db 127 EKLLLAHSWKHKNTTITRSLDILHTFASSSGFLVFNALLRAGLQWSSNAKOL 186
Qy 146 VLLKQFSNTYSNLILTEPNNTYTHMGRAVNSQGLQYIKLSLLOPRTIYRSASLYGP 205
Db 187 LGYCAQRSYNIISWELGNEFNSPKSGICIDGFLGRDFVHLRQLLSQHPLYRHAELYL 246
Qy 206 NIGRPRKNVIALLDGFMKVAGSTVDATWQHCYIDGRVVKVMDLTKRLTDLSDQIRKI 265
Db 247 DVGQPKHTQHLRFSMKGGKAIDSVTWHHYVNGRSATREDFLSPEVDSFATAIHQV 306
Qy 266 QKVNTYTPGKKIWLLEGVVTTSAGGTNNLSDSYAAGFLWMLNTLGLMANQIDVYIRHSFF 325
Db 307 LGIPEATVPGKKVWLGETSGAYGGAPQLSNTYVAGFWMLDKLGLAARSGIDVVMRQVSF 366
Qy 326 DHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVAGLQRKPRGRVTRDKLRIYAH 385
```

```
Db 367 GAGSYHLVDAGSKPLPDYWLSSLYKRLVGRVLRQASVEQADAR-RP-----RYVLHC 417
Qy 386 TTHNHNHYVRGSITLFIINLHRSRKKIKLAGTLRDKLVHQLVLPYGOGLKSKSVOLNG 445
Db 418 TNPRHPKREGDVTFLALNLSNVTQSLQPKOLWKSVDQYLLLPHGKDSILSREVQLNG 477
Qy 446 QPLVMVDGTLPELKPRLRAGRTLVIPTVMTGFFVYKVNNAAC 490
Db 478 RLLQWVDDTLPALHEMALAPGSTLGLPAFSYGFVIRNAKAIAC 522

RESULT 9
Q9SDA1
ID Q9SDA1 PRELIMINARY; PRT; 521 AA.
AC Q9SDA1;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL 57.8 KDA PROTEIN.
GN F13G24.30.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Van Der Schueren J., Chuang Y.J., Voet M., Robben J.,
RA Volckaert G., Bancroft I., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL13421; CAB62595.1; -.
DR InterPro; IPR001254; Trypsin.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 521 AA; 57831 MW; 07D8664A4B305CC2 CRC64;

Query Match 11.7%; Score 302; DB 10; Length 521;
Best Local Similarity 24.8%; Pred. No. 7.9e-16;
Matches 139; Conservative 73; Mismatches 183; Indels 166; Gaps 27;

Qy 34 NENFLSLQID--PSIIHD-----GWLDFLS---SKRLVTLA-RGLSPAFLPFGGKR 78
Db 22 DENFVCATLDWMP---HDKCNYDQCWPYSSVINMDLTRPLLTAKAKPKLRIRIGSSL 78
Qy 79 TDFLQFO--NLNPA----KSRGGPGPDYLYKNYEDDIVRSDVALDKQKCKIAQHPDVM 132
Db 79 QQVIYDVGNLKYTPCPRPFQKMNSG-----LFGFSKGCLHMKRWD-- 117
Qy 133 LELQREKAAOMHLV-----LLKEQFSNTYSNLILTEPNNTYTHMGRAVN---- 176
Db 118 -ELNSLTATGAVVTGFLNALRGRHKLRGKAWGGAMDHINTQDFLNTYTSKGVYDSWEF 176
Qy 177 GSQI-----GKDYIQLKSLLOPRTIYRSASLYGNIGRPRKNVIALLDGF- 221
Db 177 GNELSGSGVGASVSALYKGLDLTVLKDVIN--KVKNSWLHKPILVAP-----GGFY 226
Qy 222 -----MKVAG-STVDATWQHCYIDGR-----VVKVMDFLTKRLTDLSDQIRKIQ 266
Db 227 EQQWTKLEISGSPVVDVVT-HHYNLGSNDPALVKKIMD---PSYLSQVSKTFKDVN 282
Qy 267 KVVNTYTPGKKIWLLEGVVTTSAG-----GTNNLSDSYAAGFLWMLNTLGLMANQIDVYIRH 322
Db 283 QTIQEHGP---WASPWGESGGAYNSGGRHVSDDTFIDSFVWLDQLGMSARHNTKYVCRO 338
Qy 323 SFEDHCYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVAGLQRKPRGRVTRDKLRIY 382
Db 339 TLVGGFYGLLEKGTGTFPNPDYYSALLWHLRMKGVLAVQTDG-----PPQLRVY 387
```

Db 361 TLVGGFGLLEKGTFFVPNDYYSALLWHLRMGKGLAVQTGD-----PPQLRVY 409  
Qy 383 AHCTNHHNNHYVRSITLFIINLH-----RSRKKIKLAGTLRDKLV 423  
Db 410 AHCSKG-----RAGVTLLINLSNQSDFTVSVSNGINNVLNNAESRKKKSLDTLKRPFS 463  
Qy 424 -----HOYLQPYGQEG-LKSKSVOLNGQPLVMVDDGTLPKLPRLRAGRT 469  
Db 464 WIGSKASDGYLNRREYHLTP--ENGVLRSKTMVLNGSLKPTATGDIPLSLEPVLRSVNSP 521  
Qy 470 LVIPPTVMTGFFVVKVNNALAC 490  
Db 522 LNVLPMSFIVLPNFDASAC 542  
RESULT 11  
Q9FLK8 PRELIMINARY; PRT; 516 AA.  
AC Q9FLK8;  
DT 01-MAR-2001 (TremBLrel. 16, Created)  
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)  
DE SIMILARITY TO HEPARANASE.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=98290546; PubMed=9628582;  
RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.  
RT Physical features of the regions of 1,456,315 bp covered by nineteen  
RT physically assigned P1 and TAC clones."  
RL DNA Res. 5:41-54 (1998).  
DR EMBL; AB010073; BAB08480.1; -  
DR InterPro; IPR001917; AminoTransf\_2.  
DR PROSITE; PS00599; AA\_TRANSFER\_CLASS\_2; UNKNOWN\_1.  
SQ SEQUENCE 516 AA; 57414 MW; 6943E6D3D89FB93C CRC64;

Qy 383 AHCTNHHNNHYVRSITLFIINLH-----RSRKKIKLAGTLRDKLV 423  
Db 388 AHCSKG-----RAGVTLLINLSNQSDFTVSVSNGINNVLNNAESRKKKSLDTLKRPFS 441  
Qy 424 -----HOYLQPYGQEG-LKSKSVOLNGQPLVMVDDGTLPKLPRLRAGRT 469  
Db 442 WIGSKASDGYLNRREYHLTP--ENGVLRSKTMVLNGSLKPTATGDIPLSLEPVLRSVNSP 499  
Qy 470 LVIPPTVMTGFFVVKVNNALAC 490  
Db 500 LNVLPMSFIVLPNFDASAC 520  
RESULT 10  
Q9FF10 PRELIMINARY; PRT; 543 AA.  
AC Q9FF10;  
DT 01-MAR-2001 (TremBLrel. 16, Created)  
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)  
DE SIMILARITY TO HEPARANASE.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=97471969; PubMed=9330910;  
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,  
RA Miyajima N., Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence  
RT features of the 1.6 Mb regions covered by twenty physically assigned  
RT P1 clones."  
RL DNA Res. 4:215-230 (1997).  
DR EMBL; AB005249; BAB09947.1; -  
DR InterPro; IPR001254; Trypsin.  
DR PROSITE; PS00135; TRYPSIN\_SER; UNKNOWN\_1.  
SQ SEQUENCE 543 AA; 60250 MW; 0FA2248948282FF6 CRC64;  
Query Match 11.7%; Score 302; DB 10; Length 543;  
Best Local Similarity 24.8%; Pred. No. 8.4e-16;  
Matches 139; Conservative 73; Mismatches 183; Indels 166; Gaps 27;  
Qy 34 NENFLSLQLD--PSIIHD-----GWLDFLS---SKRLVTLA-RGLSPAFLEFGGR 78  
Db 44 DENEVCATLDWPP---HDKNYDQCPWGYSSVINMDLTRPLLTAKAFKPLRIRIGGSL 100  
Qy 79 TDFLOFQ--NLRNPA-----KSRGGPGDYLYKNVEDDIVRSDVALDKQCKIAQHPDVM 132  
Db 101 QDQVYIDVGNLTKPCRFQKNNG-----LFGSKCLHMKRWD-- 139  
Qy 133 LELOREKAAQHLV-----LLKEQFSNTYSNLILTEPNNYRTWHGRAVN---- 176  
Db 140 -ELNSFLTATCAVTVFTGLNALRGRHLKRGKAGGAWDHINTQDPLNTVTSKGYVIDSWEF 198  
Qy 177 GSQI-----GKDYIQKSLQPIRYSRASLYGPNIGRPRKNVIALLDGF- 221  
Db 199 GNELSGSGVGASVSAELYGKDLIVLKOVIN--KVYKNSWLHKPLVAP-----CGFY 248  
Qy 222 -----MKVAG-STVDAVTWQHCYIDGR-----VVKVMDFLKTRLLDLSQIRKI 266  
Db 249 EQQWTYKLEISGPSVVDVVT--HHIYNLGSGNDPALVKKIMD---PSYLSQVSKTFKDVN 304  
Qy 267 KVVNTYTGKKIWLKGVVTTAG-----GTNNLSDSYAAGFLWNTLGLANQIGDIVIRH 322  
Db 305 QTIQEHGP-----WASFPWVGESGAYNSGGRHSPTFIDSFWYLDOLGMSARHNTKVYCRQ 360  
Qy 323 SFFDHGYNHLVDQNFNPLPDYLSLKYKRLIGPKVLAVHVGAGLQKRPGRVIRDKLRIY 382

```
Db 322 SKHNTKCYCQALVGGFYGLLEKETFPVNPDPYISALLWHRLMGKILGVQTTA----- 375
QY 371 PGRVIRDKLRIYAHCTNHHNNHNVGRSGITLFIINLHR-----SRKK 411
Db 376 -----SEYLRAVHCSCR-----RAGIILLINLSKHTTFFVAVSNGVYVLOAESMR 424
QY 412 IKLAGFLRDKLV-----HOYLLQPYGQEG-LKSKSVQVNGOPLVMVDDGTLP 457
Db 425 KGFLEITKSKVSWGVNGKASDGYLNREYHLSR--KGDLSRKIMLNGKPLVPTATGDI 482
QY 458 ELKPRPLRAGRRLVIPPVTMGFFVKNVNALAC 490
Db 483 KLEPVRHGKSPVYINPLSISFVLPTFDAPAC 515

RESULT 12
Q9LRC8 PRELIMINARY; PRT; 527 AA.
AC Q9LRC8;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE BETA-GLUCURONIDASE.
GN SGUS.
OS Scutellaria baicalensis.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Lamiaceae; Scutellaria.
OX NCBI_TaxID=65409;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20418130; PubMed=10858442;
RA Sasaki K., Taura F., Shoyama Y., Morimoto S.;
RT "Molecular Characterization of a Novel beta-Glucuronidase from
RT Scutellaria baicalensis Georgi.";
RL J. Biol. Chem. 275:27466-27472(2000).
DR EMBL; AB040072; BAA97804.1; -
DR InterPro; IPR001179; FKBP_PPase.
DR PROSITE; PS00453; FKBP_PPase_1; UNKNOWN_1.
SQ SEQUENCE 527 AA; 58772 MW; A5DE7C423F2A1E2B CRC64;

Query Match 10.0%; Score 256.5; DB 10; Length 527;
Best Local Similarity 24.2%; Pred. No. 3.6e-12;
Matches 129; Conservative 81; Mismatches 207; Indels 115; Gaps 27;

QY 15 KEKTLILLDVSTKNPVRTVNEFLSLQD---PSIIHDG---W-----LDFLSKRLV 61
Db 26 EETIIVKIE---ENPVAQDENTVCATLDLWPPTKCYNGCPWKGKSSFLNLD-LNNNIIR 81
QY 62 TLARGLSPAFLRFGGKRTDFLOFQNLNRPAAKSRGGPGDPYLYKNYEDDIVRSDVALDKQK 121
Db 82 NAVKEFAPLKLREGGTLDRLVYQT-----SRDECDSTFYNN-----TNLIDFSH 128
QY 122 GKIAQHPDV-----LELOREKAAQMHVLKLEQ-----FSN 154
Db 129 ACLSLDRWDEINQFILETGSEAVFGLNLRGKTVEIKQIQOQYLGETTAVGEWDYSN 188
QY 155 TYSNLLITPENNYRTMHGRAVNGSQLG-----KDYI-QLKSLLQPIR-IYSRASL 202
Db 189 SKFLIEYSLUKKYKHKGWTL-GNELGGHTLFTGVSPEDYANDAKKLHELVEIKYIQDQG 247
QY 203 YGPNIGRPRKNVIALLDGFMKAVAGSVDA-VTWHQCYIDGRVVVKVDFELKTRLLD-TLSD 260
Db 248 M-PLIAP--GAIFDELTWEFTIDRTPELHVATHHMYNLSGGD--DALKDVLTTASFDD 302
QY 261 QIRK-----IQKVVNTYTPGKK--IWLEGVVTTSAGTNNLSDSYAAGFLWLNTLGLAN 313
Db 303 EATKSMYEGLOKIVN--RPTKAVAMIGEAGGAFNSGQDGISNTFFNGFWYLNWLGYSAL 360
QY 314 QGIDVIRISFFDHGYNHLVDQNFNLPDYWLSLLYKRLIGPKVLAVHVAGLQRPGR 373
```

```
Db 361 LDTKTEFCROTLLTGGNYGLLQGTGYIPNPDPYISALLWHRLMGSKVLKTEIVGTK----- 413
QY 374 VIRDKLRIYAHCTNHHNNHNVGRSGITLFIINLHRSKKIKLA-----GTLRDKLVHOYL 427
Db 414 -----NVIYIHAHCAKSN-----GITMLVLN-HGESSVKISLDPSKYSKRE-----EYH 458
QY 428 LOPYGQEGLSKSVQVNGOPLVMVDDGTLPKLPRPLRAGRRLVIPPVTMGF 479
Db 459 LTPV-NNNLQSLRVKLGELLHLLDPSGVIPALNPVEKDNKQLEVAPIYSFME 509

RESULT 13
Q9FZP1 PRELIMINARY; PRT; 536 AA.
AC Q9FZP1;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE SIMILARITY TO HEPARANASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB028613; BAB10787.1; -
SQ SEQUENCE 536 AA; 59654 MW; 24202B8E82F3DB0E CRC64;

Query Match 9.5%; Score 246; DB 10; Length 536;
Best Local Similarity 21.4%; Pred. No. 2.6e-11;
Matches 119; Conservative 86; Mismatches 195; Indels 156; Gaps 25;

QY 30 VRTVNEFLSLQD---PSIIHDG---WLDLFLSKRL-----VTLARGLSPAFLRFGG 76
Db 41 VGTIDEDFICATLDWPPKECKDYGSCSW-DHASILNLDLTIVVPVPSKNTAFAPLKRIGG 99
QY 77 KRTDFLOFQNLNRPAAKSRGGPGDPY-----YLNKTYEDDIVRSDVALDKQKCKIAQHPD 130
Db 100 TLQDIVIYET-----PDSKQPCLPFTKN-----SSILFGYTQGCPLMRWD 140
QY 131 VMLELOREKAAQMHVL-----LKE-----QFSNTYSNLIITPEPNYR----- 168
Db 141 ELNAFFRKTGKTIVFGLNALSGRSIKSNGEIGAOWNYTNAESFIRFTAENNYTIDGWELG 200
QY 169 -----TMHGRAVNGSQLGKDYIOLKSILLOPIRISRAS---LYGPNIGRPRKNVIALLD 219
Db 201 NELCGSGVGARVANQAYDITINLRIVN--RVYKNVSPMLVIGPG----- 245
QY 220 GPMKV-----AGSTVDVAVTWHQCY-----IDGRVVKVMDFLKTRLLDTSQIRKI 265
Db 246 GFPEVDWFEYLNKAENSLNATT-RHIYDLGPGVDEHLIE--KILNPSYLDQEAQSFSL 302
QY 266 QKVNTYTPGKKIWLGVVTTTSAG-----GTNNLSDSYAAGFLWLNTLGLMANGODIVIR 321
Db 303 KNIKNSSTKAVAW----VGESGGAYNSGRNLVSNAFVYSFWYLDQGLMASLYDTKYCR 358
QY 322 HSPFDHCYNHLVDQNFNLPDYWLSLLYKRLIGPKVLAVHVAGLQRPGRVIRKLR 381
Db 359 QSLIGNYGLLNTNTPNDPYISALLIWRQLMGRKALFTTFSGKT-----KIRS 407
QY 382 YAHCTNHHNNHNVGRSGITLFIINLHRSR---KKIKLAGTLRDKLVHOYLIQPY----- 431
Db 408 YTHCAQSK-----GITVLLMNLNDNTTVAKVELNNSF--SLRTHKHKMSYKRASSQL 459
QY 432 --GOEG-----LKSQSVQVNGOPLVMVDDGTLPKLPRPLRAGRRLVIPP 474
```

Db 460 FGGPNCVIOREYHHTAKDGNLHNSQTMLLNGNALQVNSMGDLPPIEPHINSTEPIITAP 519

Qy 475 VTMGFEVVKVNNALAC 490

Db 520 YSIVFVHMNVVVPAC 535

RESULT 14

Q9VE79 PRELIMINARY; PRT; 935 AA.

AC Q9VE79;

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)

DE CG14309 PROTEIN.

GN CG14309.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY.

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Mortman J.R., Vandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RA "The genome sequence of Drosophila melanogaster."

RT Science 287:2185-2195(2000).

DR EMBL; AE003721; AAF55548.1;

DR Flybase; FBgn0038611; CG14309.

SQ SEQUENCE 935 AA; 106883 MW; 10E46B721CE5DB3F CRC64;

Query Match 6.6%; Score 171; DB 5; Length 935;

Best Local Similarity 20.9%; Pred. No. 6.3e-05;

Matches 106; Conservative 76; Mismatches 163; Indels 162; Gaps 26;

Qy 18 TLILLDVSTKNPVRTVNFENFLSLQLDPSIIHDGWLDFLSRRLVTLARGLSPAFLR----- 73

Db 19 TLVLQHVN--RPYNDVNEKFVSFAVRPEDLYDA-LDGKRRKAVTNLANLLGDARIKAVNW 75

Qy 74 -----FGKRTDFLOFQNLNRPKSRGGPGPDYLYLKNYEDDI--VRSDVALDKQKCK 124

Db 76 TMIIPVPYAPDDWDMSMTLKILNTS-----YMGVITDCIWLQGTDTGTSRAK--- 122

Qy 125 IAQHPDVMLELOREKAAQMHVLVLLKEQFSNTYSNLIITEPNNYRTMHGRAVNGSQLG--K 182

Db 123 -----DIYQELRTLKL-----MTDTFKPVV-----DDWRLMGADISAGSSADETK 162

Qy 183 DYIQLKSL-----QPIRYSRASLYGPNIGRPKNVIALLDGCFMKVAGSTVDVATWQ 235

Db 163 RYVDMSKDLNTAFGTQPANMLPKSSL-----GS----- 191

Qy 236 HCYIDGRVVKVMDFLKTRLLDLSOIRKIQKVVNTYTPGKKIWLGVVVTTSAGCTNNLS 295

Db 192 YLYDSDPALRTLQOORVPLWUTLPEE-RSSQLV-----GDETDA----- 231

Qy 296 DSYAAGFLWLNTGLMANGQIDVIRHSFFDHGVNHLVD---QNFNPLPDYWLSLYLKRL 352

Db 232 -----LRWQTMGDAAASGDFVIFKRW-----NLVDFERPNS-----LYVTALEKKV 274

Qy 353 IGPKVLAHVHAGLQKRPGRVIRDKLRIYAHCTNHHNHYVRSITLFIINLHRSKKI 412

Db 275 MGRSVFPA-----RPLNAFAPSNKLYTHCA-----NAVSGGLAFMFWVNTTEQPTTI 320

Qy 413 ---KLAGTLRDKLVHOLYLPQYGOGLKSKSVOLNQPLVMVDDGTL-PELKP-----RPLR 465

Db 321 TVKTSSTSSSSSIWQVYLVLTGHQDR-----VQLNNVRLHL--NTTLRPLIKPIDPTKPLQ 372

Qy 466 AGRTLIVPVTMTGFFVVKNNVALACRY 492

Db 373 ----LITPSMAVSFWVLPDVLNLEHCQF 395

RESULT 15

082604 PRELIMINARY; PRT; 190 AA.

ID 082604;

AC 082604;

DT 01-NOV-1998 (TREMBlrel. 08, Created)

DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE T2L5.6 PROTEIN.

GN T2L5.6.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Geisel C., Smith A., Le T.;

RT "The sequence of A. thaliana T2L5";

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Washu;

RT "The A. thaliana Genome Sequencing Project.";

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Waterston R.;

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF096371; AAC62794.1;

SQ SEQUENCE 190 AA; 21212 MW; B5E356E5526007A0 CRC64;

Query Match 6.0%; Score 155; DB 10; Length 190;

Best Local Similarity 24.9%; Pred. No. 0.00011;

```
Matches 49; Conservative 32; Mismatches 70; Indels 46; Gaps 6;
QY 321 RHFDFDGYNHLVDQFNPLPDYWLSLYKRLIGPKVLAVHVAGLQKPRGVRIRDKLR 380
Db 12 RQSLIGGNYGLLNTNFTPNPDYYSALIWRQLMGRKALFTTFSGTK-----KIR 60
QY 381 IYACTNHHNHNHYVRGSITLFIINLHRSR---KKIKLAGTLRDKLVHQLQPY----- 431
Db 61 SYTHCARQSK-----GITVLLMNLDTTVAKVVELNNSF--SLRHTKHKMSYKRASSQ 112
QY 432 ---GOEG-----LKSQSVQLNGOPLVMVDDGTLPELKPRLPRLAGRTLVIP 473
Db 113 LFGGPNQVIOREEYHLLAKDGNLHSQTMLLNGNALQVNSMGDLPIPIHINSTEPTIA 172
QY 474 PVTMGFFVVKVYNALAC 490
Db 173 PYSIVFVHMRNVVVPAC 189
```

Search completed: July 30, 2002, 08:30:45  
Job time: 997 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:15:02 ; Search time 53.59 Seconds  
(without alignments)  
224.703 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_42\_534  
Perfect score: 2577  
Sequence: 1 GDRPPLPDRAGLKEKTLI.....PVTMGFFVKNVNALACRYR 493

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pap.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pap.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pap.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pap.\*
- 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pap.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	930.5	36.1	543	2	US-08-922-170B-10
2	930.5	36.1	543	4	US-09-071-739B-2
3	930.5	36.1	543	4	US-09-260-038B-2
4	926.5	36.0	543	4	US-09-181-336-13
5	883.5	34.3	532	4	US-09-181-336-15
6	805	31.2	380	4	US-09-181-336-19
7	779	30.2	380	4	US-09-181-336-17
8	94.5	3.7	685	2	US-08-878-989-1
9	94.5	3.7	685	3	US-09-136-282-2
10	94.5	3.7	685	4	US-09-272-796-1
11	94.5	3.7	685	4	US-09-505-744-2
12	93.5	3.6	934	1	US-08-215-805A-80
13	93	3.6	471	1	US-08-257-341-9
14	91.5	3.6	915	4	US-09-346-237-2
15	91.5	3.6	928	4	US-09-514-539-4
16	91	3.5	1788	2	US-08-962-284-2
17	89.5	3.5	1118	1	US-07-934-393B-2
18	89.5	3.5	1118	1	US-08-278-089A-2
19	89.5	3.5	1118	2	US-08-838-957A-2
20	89	3.5	632	4	US-09-016-080-1
21	89	3.5	1280	2	US-08-583-276-19
22	89	3.5	1280	6	5206352-4
23	88	3.4	1106	1	US-08-435-675B-5
24	88	3.4	1106	1	US-08-336-257A-8
25	87.5	3.4	1122	1	US-08-278-089A-6
26	87.5	3.4	1122	2	US-08-838-957A-6
27	87	3.4	1124	1	US-08-323-474-2

28	87	3.4	1124	5	PCT-US93-06093-2	Sequence 2, Appli
29	86.5	3.4	834	3	US-08-539-205A-6	Sequence 6, Appli
30	86	3.3	1103	1	US-08-455-543A-53	Sequence 53, Appl
31	86	3.3	1103	2	US-08-223-305C-53	Sequence 53, Appl
32	85	3.3	673	4	US-09-078-347A-2	Sequence 2, Appli
33	85	3.3	1086	6	5386025-8	Patent No. 5386025
34	85	3.3	1280	2	US-08-752-447-2	Sequence 2, Appli
35	84.5	3.3	434	2	US-08-989-925-1	Sequence 1, Appli
36	84.5	3.3	484	3	US-09-080-044-7	Sequence 7, Appli
37	84.5	3.3	533	1	US-08-220-151-15	Sequence 15, Appl
38	84.5	3.3	533	1	US-08-413-118-15	Sequence 15, Appl
39	84.5	3.3	533	3	US-08-473-446-15	Sequence 15, Appl
40	84.5	3.3	1279	2	US-08-784-649A-2	Sequence 2, Appli
41	84	3.3	538	4	US-09-175-928-4	Sequence 4, Appli
42	84	3.3	1056	2	US-08-627-873-7	Sequence 7, Appli
43	83.5	3.2	1507	6	5268270-2	Patent No. 5268270
44	83	3.2	490	4	US-08-740-223A-12	Sequence 12, Appl
45	83	3.2	497	4	US-08-740-223A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-08-922-170B-10  
; Sequence 10, Application US/08922170B  
; Patent No. 5968822  
; GENERAL INFORMATION:  
; APPLICANT: Iris Pecker, Israel Vlodavsky and Elena  
; APPLICANT: Feinstein  
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A POLYPEPTIDE  
; TITLE OF INVENTION: HAVING HEPARANASE ACTIVITY AND EXPRESSION OF  
; NUMBER OF INVENTIONS: SAME IN TRANSDUCED CELLS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein  
; STREET: 2940 Birchtree lane  
; CITY: Silver Spring  
; STATE: Maryland  
; COUNTRY: United States of America  
; ZIP: 20906  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk  
; COMPUTER: Twinhead\* Slimnote-890TX  
; OPERATING SYSTEM: MS DOS version 6.2,  
; OPERATING SYSTEM: Windows version 3.11  
; SOFTWARE: Word for Windows version 2.0 converted to  
; SOFTWARE: an ASCII file  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/922,170B  
; FILING DATE: 2 SEP 1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Friedman, Mark M.  
; REGISTRATION NUMBER: 33,883  
; REFERENCE/DOCKET NUMBER: 910/1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 972-3-5625553  
; TELEFAX: 972-3-5625554  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-922-170B-10

Query Match 36.1%; Score 930.5; DB 2; Length 543;

```

, FILING DATE: September 2, 1997
, APPLICATION NUMBER:
, FILING DATE:
, ATTORNEY/AGENT INFORMATION:
, NAME: Friedmam, Mark M.
, REGISTRATION NUMBER: 33,883
, REFERENCE/DOCKET NUMBER: 910/5
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: 972-3-5625553
, TELEFAX: 972-3-5625554
, TELEX:
, INFORMATION FOR SEQ ID NO: 2:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 543
, TYPE: amino acid
, STRANDEDNESS: single
, TOPOLOGY: linear
, JS-09-071-739B-2

```

Query Match	Score 930.5;	DB 4;	Length 543;
Best Local Similarity	36.1%;		
Pred. NO. 8.5e-90;	38.7%;		
Mismatches 170;			
Conservative	78.		
Gaps			
Indels	89;		

Qy	1	GDRRLPVDRAGLUKEKTLIILDDYSTKNPVTNENFSLQDSDSIHD-CWLDFLSSKR	59
Db	22	GPLSPGALPRPA--QAQDVDFDFTQEBPLHVSFSLVTDANLADPRFLILGSPK	79
Qy	60	LVTARGLSPAFLRRGGKRTDFLOPNLRNPAKSRGGPDYIYLNKYEDDITVRSDVALDK	119
Db	80	LRTLARGUSPAYLRGGGKTDLIF---DPKKESTFEERSYQWQOVQDI-----	126
Qy	120	QKGCKIAQ-HPDVMLELQREKAAQHLVLKKEQF-----SNTYS-----	157
Db	127	---CKYGSIPDPVEKLRLEWPYQEQ-LLRHYQKFKNSTYSRSSVDVLYTFANCQGL	182
Qy	158	-----NUL-----TEPNNYRTMHGRAVNGSOL	180
Db	183	DLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYINISWELGNEPSFLKADIFNGSOL	242
Qy	181	KDXYTQLKSLLOPIRIYSRASLYGNICRPRKNVIALLDGFMKVAGSVTDVATWQHCYID	240
Db	243	GEDYIQLHKLRRK-STFNKAKLYGDVCGPRKTKAKMLKSFLKAGGEVIDSVTWHYYLN	301
Qy	241	GRWVKYMDPLKTRLLDTSQIRKIQKVNVYTPGKKIWLGVVTSAGGTNNLSDSYAA	300
Db	302	GRATREDFLNDVDLIDFISSSVQKVFQVVESTRPGKKVWLGETSAYYGGGAPLLSDTFAA	361
Qy	301	GFELWTLGMLANQOIDDVIRHSFDPDHGYNHLYVDQNFNPDLDYWLSLKRLIGPKVKLAV	360
Db	362	GFWMULDGLSARMGIEVYMRQVFFGAGNYHLVDENFDPDLDYWLSDLPLKVLGVTKYVMA	421
Qy	361	HVAGLQKRPGRVIRDKLRIYAHCTNHNHNHVVRGSITFLTIINLHRSRKKIKLAGTLRD	420
Db	422	SVQGSRR-----KLRYLHCTINTONPRYKEGDLTYALNHNVTYKRLPYPEFN	472
Qy	421	KLHVQYVLLPYQOEGLSKSVQLNGOPLYVMDDGTLPCLKRPLRPLRAGRTLVIPIPTMGFF	480
Db	473	KQVDKYLLRPLRPHGLLSKSVQLNGLTLMKWDDQTLUPLMEXPLRPGSSLGLPAFSYSEF	532
Qy	481	VKNVNALAC	490
Db	533	VIRNAKVAAC	542

```

RESULT 3
US-09-260-038B-2
; Sequence 2, Application US/09260038B
; Patent No. 6348344
; GENERAL INFORMATION:
; APPLICANT: Maty Ayal-Hershkovitz et al.
; TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR
; EXPRESSING RECOMBINANT HEPARANASE
;

```

```

Best Local Similarity 38.7%; pred. No. 8.5e-90;
Matches 213; Conservative 76; Mismatches 170; Indels 89; Gaps 11;

Qy      1  GDRPLPVDRAAGLUKEKTLILLDVSTKNPVTVENFLSLQLDPSIIHD-CWLDFLSSKR 59
Db      1  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy     22  GPLSPGALPRPA--QAQDVLDLDFEFTQPLHLVSPFSVLTIDANLATDPRFLILLGSPK 79
Db     22  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy     60  LVTIARGLSPAFLRFEGKRTDFLOQNLRNPAKSRGGPDYILKNYEDDITVRSDVALDK 119
Db     60  LVTIARGLSPAFLRFEGKRTDFLOQNLRNPAKSRGGPDYILKNYEDDITVRSDVALDK 119
Qy     80  LRTLARGLSPAYLREGGKTDFLIF---DPKKESTFEERSYQSQVNQDI----- 126
Db     80  LRTLARGLSPAYLREGGKTDFLIF---DPKKESTFEERSYQSQVNQDI----- 126
Qy    120  QKGCKIAO-HPDVMLELOREKAAQHVLVLLKEQF-----SNTYS----- 157
Db    120  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    127  ---CKYSGIPDPVEEKLREWPYEQOL--LLREHYOKKPKNSTYSKSSVDVLYTFANCSSL 182
Db    127  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    158  -----NLIL-----TPNNYRTMHGRAVNGSQL 180
Db    158  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    193  DLIFGLNALLRTADLQWNSSNAQLLLDYCSSKGYNIWELGNEPNSFLKKADIFINGSQL 242
Db    193  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    191  GKDYIQLKSLQLQPIRIYRSASLYGNIGRPNKNVTALLDGMKAVGSVDTVWQHCYID 240
Db    191  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    243  GEDYIQLHKLRL--STFNKAKYLPDVGQPRKTKAKLKSFUKAGEVDSIVTWHYYILN 301
Db    243  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    241  GRVVKVMDFLKTRLEDLTLSQIRKTKQKVNTYTPGKKIWLGVVTSAGTNNLSDSYAA 300
Db    241  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    302  GRTATREDFLNPVDLIDIFSSQKVFQVVESTPRCKKVMWLGTSAYGGGAPLLSDTEAA 361
Db    302  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    301  GFLWNTLGMLANQGDIVVIRHSFDHGYNHLVDQNFNPLPDYWLSLLYKRLIGPKVLAV 360
Db    301  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    362  GPMWLDKLGLSARMGIEVVMQVFFGAGNYHLVDENFDPLPDYWLSDLFPKLVGKVLMA 421
Db    362  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    361  HVAGLQRKPRGRVIRDKLRIYAHCTNHNHNHNHVVRGSIPLFTIINLHRSRKKIKLAGTLD 420
Db    361  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    422  SVQGSKRR-----KLRVYLHCTNTDNPYKEGDLTYALNHNVTLYKRLPYPSFN 472
Db    422  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    421  KLVHQYVLQPYQCEGLKSKSVQINQGPLVWVDGTLPELKPRPLRAGRTVLIVPPYTMGFF 480
Db    421  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    473  KQVDYLLRPLPHGLLSKSVQNLGLTLKMWDDQTLPLPMEKPLRPGRSSILGUPAFYSYFF 532
Db    473  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    481  VYKNYNALAC 490
Db    481  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    533  VTRNAKVAAC 542
Db    533  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT<sup>2</sup>  
US-09-071-739B-2  
; Sequence 2, Application US/09071739B  
; Patent No. 6177545  
; GENERAL INFORMATION:  
; APPLICANT: Iris Pecker et al.  
; TITLE OF INVENTION: HEPARANASE SPECIFIC MOLECULAR PROBES  
; NUMBER OF INVENTION: AND THEIR USE IN RESEARCH AND MEDICAL  
; NUMBER OF INVENTION: APPLICATIONS  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina  
; STREET: 20001 Jefferson Davis Highway, Suite 207  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: United States of America  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk  
; COMPUTER: Twinhead\* Slimnote-890TX  
; OPERATING SYSTEM: MS DOS version 6.2,  
; OPERATING SYSTEM: Windows version 3.11  
; SOFTWARE: Word for Windows version 2.0 converted to  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071.739B  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIORITY DATA:  
; PRIOR APPLICATION NUMBER: 08/922,180



Db 362 GFMWLDKLGLSARMGIEVVMRQVFFGAGNVLVDENFDPLDPLDYWLSLFLFKLVGTVKLMA 421  
 QY 361 HVAGLQKRPGRVIRDKLRIYAHCTNNHNNYVRSITLFIINLHRSRKKIKLAGTLRD 420  
 Db 422 SVQGSRR-----KLRVYLHCTNDPNRYKEGDTLTAIINLHNVTKYLRPLPFSN 472  
 QY 421 KLVHXYLLQPYGQGLKSKSVQNLGOLPLVMVDDGTLPKLPRLRAGRTL 470  
 Db 473 KQVDKYLRLPLGPHGLLSKSVQNLGLTLKMWDDQTLPLMEKPLRPGSSL 522

RESULT 6

US-09-181-336-19  
 ; Sequence 19, Application US/09181336A  
 ; Patent No. 6242238  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FREEMAN, Craig Geoffrey  
 ; APPLICANT: PARISH, Richard  
 ; APPLICANT: HAMDORF, Brenton James  
 ; APPLICANT: HULETT, Mark Darren  
 ; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN  
 ; TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR  
 ; FILE REFERENCE: 032505-004  
 ; CURRENT APPLICATION NUMBER: US/09/181,336A  
 ; CURRENT FILING DATE: 1998-10-28  
 ; EARLIER FILING DATE: 1997-10-28  
 ; EARLIER APPLICATION NUMBER: AU PP0062  
 ; EARLIER FILING DATE: 1997-12-09  
 ; NUMBER OF SEQ ID NOS: 23  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 19  
 ; LENGTH: 380  
 ; TYPE: PRT  
 ; ORGANISM: Rattus sp.  
 ; US-09-181-336-19

Query Match 31.2%; Score 805; DB 4; Length 380;  
 Best Local Similarity 47.3%; Pred. No. 9.5e-77;  
 Matches 155; Conservative 57; Mismatches 106; Indels 10; Gaps 2;

QY 163 EPNYRTHMGRVAVGSQLGKDYIQLKSLLOPIRIYSRASLYGPNIGRPNKNNVIALLDGFM 222  
 Db 62 EPSEWFKAHISIDGLQGEDFVELHLLQK-SAFQNAKLYGPDIGQPGKTVKLLRSFL 120  
 QY 223 KVAGSTVDVAVTWHQYIDGRVVKVMDFLKTRLLDTSQIRKIQKVNTYTPGKKIWLEJ 282  
 Db 121 KAGGEVDSLTWHYLLNGRVATKEDFLSSDVLDFTILSVQKILKVTREMPGKKWILGE 180  
 QY 283 VYTTAGGTNNLSDSYAAAGFLMLNTLGMLANOGIDVIRHFFDHGYNHLYVDONENPLPD 342  
 Db 181 TSSAYGGGAPLLSDTFAAGFMWLDKLGLSAQGLIEVVMRQVFFGAGNVLVDENFEPLPD 240  
 QY 343 YWLSLLYKRLGPKVLAVHAGLQKRPGRVIRDKLRIYAHCTNNHNNYVRSITLFI 402  
 Db 241 YWLSLLYKRLGPKVLAVHAGLQKRPGRVIRDKLRIYAHCTNNHNNYVRSITLFI 291  
 QY 403 INLHRSRKKIKLAGTLRDKLVHXYLLQPYGQGLKSKSVQNLGOLPLVMVDDGTLPKLPRL 462  
 Db 292 LNLHNVTKYLRPLPFSN-----KLRVYLHCTNDPNRYKEGDTLTAIINLHNVTKYLRPLPFSN 472  
 QY 463 PLRAGRTLVIPIVPMGFFVVMKNNALAC 490  
 Db 352 PLPAGSSLSVPAFSYGFVIRNAKIAAC 379

RESULT 7

US-09-181-336-17  
 ; Sequence 17, Application US/09181336A  
 ; Patent No. 6242238  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FREEMAN, Craig Geoffrey

Db 362 GFMWLDKLGLSARMGIEVVMRQVFFGAGNVLVDENFDPLDPLDYWLSLFLFKLVGTVKLMA 421  
 QY 361 HVAGLQKRPGRVIRDKLRIYAHCTNNHNNYVRSITLFIINLHRSRKKIKLAGTLRD 420  
 Db 422 SVQGSRR-----KLRVYLHCTNDPNRYKEGDTLTAIINLHNVTKYLRPLPFSN 472  
 QY 421 KLVHXYLLQPYGQGLKSKSVQNLGOLPLVMVDDGTLPKLPRLRAGRTL 470  
 Db 473 KQVDKYLRLPLGPHGLLSKSVQNLGLTLKMWDDQTLPLMEKPLRPGSSLGPAFSYFF 532  
 QY 481 VYKNNVALAC 490  
 Db 533 VIRNAKVAAC 542

RESULT 5

US-09-181-336-15  
 ; Sequence 15, Application US/09181336A  
 ; Patent No. 6242238  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FREEMAN, Craig Geoffrey  
 ; APPLICANT: PARISH, Richard  
 ; APPLICANT: HAMDORF, Brenton James  
 ; APPLICANT: HULETT, Mark Darren  
 ; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN  
 ; TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR  
 ; FILE REFERENCE: 032505-004  
 ; CURRENT APPLICATION NUMBER: US/09/181,336A  
 ; CURRENT FILING DATE: 1998-10-28  
 ; EARLIER FILING DATE: 1997-10-28  
 ; EARLIER APPLICATION NUMBER: AU PP0062  
 ; EARLIER FILING DATE: 1997-12-09  
 ; NUMBER OF SEQ ID NOS: 23  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 15  
 ; LENGTH: 532  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-181-336-15

Query Match 34.3%; Score 883.5; DB 4; Length 532;  
 Best Local Similarity 38.7%; Pred. No. 7.9e-85;  
 Matches 205; Conservative 75; Mismatches 161; Indels 89; Gaps 11;

QY 1 GDRPLPVDRAAGLKEKTLILLDVSTKNPVRTVYNNFLSLQDPSIHD-GWLDLSSKR 59  
 Db 22 GLSPGALRPA--QAQDVVDLDFTEQPLHLVSPSFLSVTIDANLATDPRFLLIGSPK 79  
 QY 60 LVTLARGLSPAFLEFGKRTDFLQFNLRNPAKSRGGPGDPDYLLKNVEDDIVRSDVALDK 119  
 Db 80 LRTLARGLSPAFLEFGKRTDFLQFNLRNPAKSRGGPGDPDYLLKNVEDDIVRSDVALDK 119  
 QY 120 QKGCIAQ-HPDVMLELQREKAAQMLHLVLEKEF-----SNTYS----- 157  
 Db 127 --CKYGSIPDVEEKLRLWMPYQEQ-LLREHYQKKFKNSTYSRSVDVLYTFANCSSL 182  
 QY 158 -----NLIL-----TEPNYRTHMGRVAVGSQL 180  
 Db 183 DLIFGLNALLFTADLOWNSSNAQLLLDYCSKGYNTISWELGNFNSFLKKADIFINGSQL 242  
 QY 181 GKDYIQLKSLLOPIRIYSRASLYGPNIGRPNKNNVIALLDGFMKRVAGSTVDVAVTWHQYID 240  
 Db 243 GEDFIQLHKLRLK-STFNKAKLYGPDVGPQPRRTAKMLKSLKAGGEVIDSVTWHYLLN 301  
 QY 241 GRVVKVMDFLKTRLLDTSQIRKIQKVNTYTPGKKIWLEGVVTSAGGTNNLSDSYAA 300  
 Db 302 GRATREDFLNPVDLIDFISVQVQVFWESTRPGKKWILGETSSAYGGGAPLLSDTFAA 361  
 QY 301 GFLWNTLGMLANOGIDVIRHFFDHGYNHLYVDONENPLPDYWLSLFLFKLVGTVKLAV 360

APPLICANT: PARISH, Richard  
APPLICANT: HAMDRE, Brenton James  
APPLICANT: HULETT, Mark Darren  
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN  
TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR  
FILE REFERENCE: 032505-004  
CURRENT APPLICATION NUMBER: US/09/181.336A  
CURRENT FILING DATE: 1998-10-28  
EARLIER APPLICATION NUMBER: AU PP0062  
EARLIER FILING DATE: 1997-10-28  
EARLIER APPLICATION NUMBER: AU PP0812  
EARLIER FILING DATE: 1997-12-09  
NUMBER OF SEQ ID NOS: 23  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 17  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-181-336-17

Query Match 30.2%; Score 779; DB 4; Length 380;  
Best Local Similarity 46.3%; Pred. No. 5.4e-74;  
Matches 152; Conservative 56; Mismatches 110; Indels 10; Gaps 2;

Qy 163 EPNYRTHMGRVNGSQGKDYIQLSKLLQPIRYSRASLYGNIPGRKVNIALLDGFM 222  
Db 62 EPNFHWKAHLIDGLQGEFVELHKLQR-SAFONAKLYGPDIGQPRGKTVKLLRSFL 120  
Qy 223 KVAGSTVDATWQHCYIDGRVVKVWDFLKLRLDLSQIRKIQKVNTYTPGKKINLEG 282  
Db 121 KAGGEVDLSLTHHHYLNGRATDEDFLSSVDLTFTILSVOKILKVTREITPGKKVWLGE 180  
Qy 283 VVTSAGTNNLSYSAAGFLWLNTLGMANOGIDVIRHSHFFDHGYNHLVDONFNPLPD 342  
Db 181 TSSAYGGGAPLLSNTFAGFWLKLGLSQAQMGIEVNRQVFFGAGNTHLVDENEPEPLPD 240  
Qy 343 YWLSLLYKRLGPKVLAVHVGAGLQKRPGRVIRDKLRIYAHCTNHHNHNHYVRSITLFI 402  
Db 241 YWLSLLEKLVGPRVLLSRVAGPD-----RSKLVYLHCTNVYHPRYQEGDLTLVY 291  
Qy 403 INLHRSRKIKLAGTLRDKLVHQVLLQPYGOEGLKSKSVOLNGQPLVMVDGTLPELKPR 462  
Db 292 LNLHNVTKHLVPPPLFRKPVDTYLLKPSGPDGLLSKSVQLNGQLKMWDEQTLPLATEK 351  
Qy 463 PLRAGRTLVIPPTVMGFFVWKNVNALAC 490  
Db 352 PLPAGSALSPLAFSGFVIRDAKIAAC 379

RESULT 8  
US-08-878-989-1  
Sequence 1, Application US/08878989  
Patent No. 5885803  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
APPLICANT: Guegler, Karl G.  
APPLICANT: Lal, Preeti  
APPLICANT: Goli, Surya K.  
APPLICANT: Shah, Purvi  
TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN  
TITLE OF INVENTION: KINASES  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/878,989  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0321 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 685 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: HUVENOB01  
CLONE: 39043  
US-08-878-989-1

Query Match 3.7%; Score 94.5; DB 2; Length 685;  
Best Local Similarity 23.1%; Pred. No. 0.52;  
Matches 84; Conservative 46; Mismatches 144; Indels 89; Gaps 18;

Qy 88 RNPASRGPGPDYLLKNYEDDIVRSVALDKQCKTAQHPDVMLELQREKAAQ----- 142  
Db 84 RGVKLGKG-----PAKCYE-----MTDLTNKVAAKIIPHSRVAKPHQREKIDKEIELH 134  
Qy 143 -----MHLVLLKEQFSNTYSNLIITPEPNYRTMHGRAVNSQLGKDYIQLSKLLQPIRI 196  
Db 135 RILHHKVVQVHYHFDKENIYILLEYCSRSM-----AHILK-----ARKVLTPEVRY 184  
Qy 197 YSRASLYGNIPGRKVNIALIDGFMKVGAGSTVDATWQHCYIDGRVVKVMDFLKTRLLD 256  
Db 185 YLRQIVSGLKYLHEQE-----ILHRLDKLGNEFIN-----EAMELKVGDFGLAARLE 231  
Qy 257 TLDQIRKIQKVNTYTP-----GKXWLEGVV--TTSAG-----GTNNLSDSY-- 298  
Db 232 PLEHRRRTICGTPNYLSPEVLNKQHGCESDIWLGCVMYTMLLGRPPPFETTNLKETYRC 291  
Qy 299 --AAGFLWLNTL-----GMLANQ-----GIDVIRHSHFFDHGY--NHLVDONFNPL 340  
Db 292 IREARYTPWPSLLAPAKHLIASMSJKNPDRPSLDDIIRHDFFLQGFDPDRLSSSCCHTV 351  
Qy 341 PDYWLSLLYKRLGPKVLAVHVGAGLQKRP-----PGRVIRKRLRIYAHCTNHHNHNHYVRG 396  
Db 352 PDFHLLSPAKNFF-KAAAAALFGGKKDKARYIDTNHRSKEDEDIY-----KLRLDLAKT 405  
Qy 397 SIT 399  
Db 406 SIT 408

RESULT 9  
US-09-136-282-2  
Sequence 2, Application US/09136282  
Patent No. 6063609  
GENERAL INFORMATION:  
APPLICANT: ANDERSON, KAREN  
APPLICANT: JACKSON, JEFFREY  
APPLICANT: HANSBURY, MICHAEL  
APPLICANT: NERURKAR, SANDHYA

	Query Match	3.78;	Score 94.5;	DB 3;	Length 685;
	Best local Similarity	23.1%;	Pred. No. 0.52;	144;	Indels 89;
	Matches 84;	Conservative			Gaps 18;
Qy	88	RNPASRGGPGDYYLKNYEDDIIVRS	VDALDGKQCKIAQHPDVNMLELQREKAAQ	-----	142
Db	84	RKVLGRGQ	-----FAKCYE	---MTDLTNKKVYAAKIIPHISRVAKPHQREKIDKEIELH	134
Qy	143	-----MHVLVLKEQFSNTYSNLIIT	EPNNYRTMHGRAVNGSQLGDKDYITQLKSLLOP	-TRI	196
Db	135	RI LHKKHVQYHYFEDKENIYILLEYS	CRSRM-----AHILK	---ARKVLTEPEVRY	184
Qy	197	YSRASLVGPNIGRPRKNVIALDGF	GMK VAGSTVDAVTWQHCVYIDGRVVKVMDFLKRLLD	256	
Db	195	YLRQIVSGLKYLHEQ	-----ILHRDLKLGNFFIN	-----EAMELKVGDFGLAARLE	231
Qy	257	TLSDQIRKIQKVNTYTP	-----GKKIWLEGVV	-TTSAG-----GTNNLSDSY	-- 298
Db	232	PLEHRRRTICGTPNYLSPEVLNKQ	HGHCESDIWALGCVNYTMLLGRPPFETTNLKETYRC	291	
Qy	299	--AAGFLWLNTL	-----GWLANQ	-----GIDVIRHSFEDHG	-- NHLVDQNFPL 340
Db	292	IREARTYMPSSLLAPAKHLIASML	SKNPNEDRPSLDIILRHDFE	LOGFTPDRLSSSCCHTV	351
Qy	341	PDYWLSSLYLKRIGPKVLAVH	VAGLQRKPR	-----PGRVIRDKLRTIAHCTHNHHNNHYR	396
Db	352	PDFHLSSPAKNFF	-KKAAMALFGGKKDKARYID	TDHNRSKEDEDIY	-----KLRLDLKKT 405
Qy	397	SIT 399			
Db	406	SIT 408			

	Query Match	3.7%	Score 94.5	DB 4	Length 685	
	Best Local Similarity	23.1%	Pred. No. 0.52			
	Matches	84	Conservative	46	Mismatches	144
					Indels	89
					Gaps	18
QY	88	RNPAKSRGGPGDYILKNYEDDIIVSRDVALDKQKCKIAQHPDVMLELQREKAAQ	-----	142		
Db	84	RGKVLGKGG	-----FAKCYE-----MTDLTNKVVAAKIIIPHSRVAKPHQREKIDETELH	134		
QY	143	-----MHLVLLKKEQFSNTYSLNLTPENNVYRTMHGRAVNGSOLCKDVIQLKSLQP-IRI	196			
Db	135	RILHKKHVQFYHFDEKENIYILLEVCSSRM	-----AHILK-----ARKVLTEPEVRY	184		
QY	197	YSRASLYGPNIGRPRKNVIALLDGFMKVGSTVDVATWQHGYIDGRVVKVWDFLKRLLD	256			
Db	185	YLROIVSGLKYLHQE	-----ILHRDLKLGNNFFIN-----BAMELUKVGDFGLGAARLE	231		
QY	257	TLSQDIIRIKIVQVNTYTP	-----GKKIWLEGVY--TTSAG-----GTNNJSDSY--	298		

Db 232 PLEHRRRTICGTPNYLSPEVLNKGHGCESDIWALGCVNMTMLGRPPFTTNLAKETRC 291

Qy 299 --AAGFLWLNTL-----GMLANQ-----GIDVIRHSHFFDHGY--NHLVDQNFNPL 340

Db 292 IREARYTMPSSLLAPAKHLIASMLSKNEDRPSLDDIIRHDFFLQGFPPDRLSSSCCHTV 351

Qy 341 PDYWLSLLYKRLIGPKVLAVHAGLQRPKPR-----PCRVIRDKLRIYAHCTNHHNNHNYVRG 396

Db 352 PDFHLSSPAKNFF-KKAAALFGGKKKARYIDTHNRVSKEDEDIY-----KLKRDLLKKT 405

Qy 397 SIT 399

Db 406 SIT 408

RESULT 11

US-09-505-744-2

; Sequence 2, Application US/09505744

; Patent No. 6245544

; GENERAL INFORMATION:

; APPLICANT: Karen M. Anderson

; APPLICANT: Mark M. Bouzyk

; APPLICANT: Michael J. Hansbury

; APPLICANT: Jeffrey R. Jackson

; APPLICANT: Sandhya S. Nerurkar

; APPLICANT: Amy K. Roshak

; TITLE OF INVENTION: HUMAN SERUM INDUCIBLE KINASE (SNK)

; FILE REFERENCE: GH-70231-D1

; CURRENT APPLICATION NUMBER: US/09/505,744

; CURRENT FILING DATE: 2000-02-16

; EARLIER APPLICATION NUMBER: 09/136,282

; EARLIER FILING DATE: 1998-08-20

; EARLIER APPLICATION NUMBER: 60/056,112

; EARLIER FILING DATE: 1997-08-20

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 685

; TYPE: PRT

; ORGANISM: HOMO SAPIENS

US-09-505-744-2

Query Match 3.7%; Score 94.5; DB 4; Length 685;

Best Local Similarity 23.1%; Pred. No. 0.52;

Matches 84; Conservative 46; Mismatches 144; Indels 89; Gaps 18;

Qy 88 RNPASRGSGPGDYLLKNYEDDIVRSVDALDKQCKIAQHPDVMLELQREKAAQ----- 142

Db 84 RGVLRGKG-----FAKYE-----MTDLTNKKVYAAKIIPHSRVAKPHQREKIDKEIELH 134

Qy 143 -----MHLVLLKEQFSNTYSNLIITPEPNNTMHRVNGSOLGKDYIQLKSLLOP-IRI 196

Db 135 RLTHKHVVQVHYFEDKENIYILLECYSSRM-----AHILK---ARKVLTEPEVRY 184

Qy 197 YSRASLYGPNTRPKKNYIALDGPMPKAVAGTVDAVTWQHCVYIDGRVYKVMDFLKLRLD 256

Db 185 YLRQVSLGKYLHEQE-----ILHRDLKLGNFN-----EAMELKVGDFGLAARLE 231

Qy 257 TILSDQIRKIQKVVNTYTP-----GKKIWLGVV--TTSAG-----GTNNLSDSY-- 298

Db 232 PLEHRRRTICGTPNYLSPEVLNKGHGCESDIWALGCVNMTMLGRPPFTTNLAKETRC 291

Qy 299 --AAGFLWLNTL-----GMLANQ-----GIDVIRHSHFFDHGY--NHLVDQNFNPL 340

Db 292 IREARYTMPSSLLAPAKHLIASMLSKNEDRPSLDDIIRHDFFLQGFPPDRLSSSCCHTV 351

Qy 341 PDYWLSLLYKRLIGPKVLAVHAGLQRPKPR-----PCRVIRDKLRIYAHCTNHHNNHNYVRG 396

Db 352 PDFHLSSPAKNFF-KKAAALFGGKKKARYIDTHNRVSKEDEDIY-----KLKRDLLKKT 405

Qy 397 SIT 399

Db 406 SIT 408

Db 406 SIT 408

RESULT 12

US-08-215-805A-80

; Sequence 80, Application US/08215805A

; Patent No. 5559008

; GENERAL INFORMATION:

; APPLICANT: Chang, Yung-Fu

; TITLE OF INVENTION: LEUKOTOXIN GENE FROM PASTEURELLA

; TITLE OF INVENTION: SUIS

; NUMBER OF SEQUENCES: 84

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Nixon, Hargrave, Devans & Doyle

; STREET: Clinton Square, P.O. Box 1051

; CITY: Rochester

; STATE: New York

; COUNTRY: USA

; ZIP: 14603

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC Compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/215.805A

; FILING DATE: 22-MAR-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Timian, Susan J.

; REGISTRATION NUMBER: 34,103

; REFERENCE/DOCKET NUMBER: 19603/61 (D-1329A)

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (716) 263-1636

; TELEFAX: (716) 263-1600

; INFORMATION FOR SEQ ID NO: 80:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 934 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Pasteurella suis

; STRAIN: 5943

; IMMEDIATE SOURCE:

; LIBRARY: P. suis DNA in Bacteriophage lambda-dash

; CLONE: (Lambda)yfc33-37

US-08-215-805A-80

Query Match 3.6%; Score 93.5; DB 1; Length 934;

Best Local Similarity 24.3%; Pred. No. 1.1;

Matches 64; Conservative 30; Mismatches 80; Indels 89; Gaps 13;

Qy 101 YLKNYEDDIVRSVDALDKQCKIAQHPDVMLELQREK-----AAOMHLVLLKEQFSNT 155

Db 43 YIPKDYEDSGRNGQLDLVAAE-----DLGIEVQREERNGIATAQNSLSTIQTILGFS 97

Qy 156 YSNLIITPEPNNTMHRVNGSOLGKDYIQLKSLLOPIRIYSRASLYGPNTRPKKNY 215

Db 98 ERGVVLSAP-----QLDKLLQYKI-SKAPGSSENVAKNLNAQ 135

Qy 216 ALLDGFMRVAGSTVDVATWQHCVYIDGRVVK-----VMDFLKTRL-----L 255

Db 136 TLLSGIQLSGVMAQMD-----LD-ETLKNKGSELDLAKAGLELTNSLIENIANSVQTL 189

Qy 256 DTLSQDI-----RKIQKVVNTYTPGKK-----IWLEGVWTTAGGTNN--LSDSY 298

Db 190 DTFSEQISQLGTKLQNVKGLGTGLDKLNFSGFSKAGLGLEVISGLSGATAALVLADKN 249

Qy 299 A-----AGFLWLNTLGLMANQ 314

Db 250 ASTDRKVGAGF-----ELANQ 265

RESULT 13  
US-08-257-341-9  
; Sequence 9, Application US/08257341  
; Patent No. 5525491  
; GENERAL INFORMATION:  
; APPLICANT: HUSTON, JAMES S  
; APPLICANT: OPPERMANN, HERMANN  
; APPLICANT: TIMASHEFF, SERGE N  
; TITLE OF INVENTION: SERINE RICH PEPTIDE LINKER  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CREATIVE BIOMOLECULES, INC./PATENT DEPT.  
; STREET: 35 SOUTH STREET  
; CITY: HOPKINTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 01748  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/257,341  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/842,149  
; FILING DATE:  
; APPLICATION NUMBER: US 07/662,226  
; FILING DATE: 27-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CAMPBELL ESO, PAULA A  
; REGISTRATION NUMBER: 32,503  
; REFERENCE/DOCKET NUMBER: CRP-064CP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/248-7000 (ATTY)  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 471 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-257-341-9

Query Match 3.6%; Score 93; DB 1; Length 471;  
Best Local Similarity 19.2%; Pred. No. 0.4;  
Matches 90; Conservative 71; Mismatches 152; Indels 156; Gaps 21;

QY 14 LKEKTLILLDVSTKPNVTVNENFLSLQDPSIIHDGWLDFLSSKRLVTLARGLS----- 68  
Db 68 LADRINIVLSRELKEPPR--GAHFLAKSLDDALRIEQPELASKYDMVMIVGGSSVYQEA 125  
QY 69 ---PAFLRFGGKRTDFLOFQNLNRPAPKSRGGPGPDYILKNYEDDIIVRSDVALDKOGCKI 125  
Db 126 MNQPGHLRLFVTR-----INQEFESDFFFEIDLKGYK--LL 160  
QY 126 AQHPDVMLEQREKAAQMHVLKLEQFSNTYSNLITPEPNRYTMHGRAVNSQLGKDYI 185  
Db 161 PEYPGVLSEVOEKGKIKYKFEYERKKDAS-----VTSSSGSSSSSSSGSSGSMIEQD-- 212  
QY 186 QLSKLLQPIRIYSRASLYSGNIGRPNKVNIAL--DGFPMKAGSVTDVATWQHCHYIDRV 243  
Db 213 -----GLHAGSPAANVERLFGVDWAQQITGCS--DAAVFR--LSAQGRP 252  
QY 244 VKVMDFLKRLLDLSQDQIRKIQKVVNTYTPGKKIWLKLEGVVTTTSAGGTNNLSDSYAAGFL 303  
Db 253 VL---FVKTDL-----SGALNELQDE--AARLS 275

QY 304 WLNTLGLMLANOGIDVYVIRHFFDHGYNHLVDQNFENPLPDYWLSLLYKRLIGPKVLAVHVA 363  
Db 276 WLATTGVCPCAAVLDVVT-----EAGRD-----W--LLLGEVPGQDLSSHLA 315  
QY 364 GLQRKPRPGRVTRDKLRIY-----AHCTNNHHNNHNVVRSITLFIINLHRSRKKIKLAGTL 418  
Db 316 PAEKV-----SIMADAMRRRLHTLDPATCPDHOAKH-----RIERARTRMEAGLVD 361  
QY 419 RDKLVHXYLLQPYGOEGL-----KSKSVQLNGQPLVMV-DDGTLP 459  
Db 362 QDDLEEH-----OGLAPAEFLARLKARMPDGEDLVVTHGDACLPNI 403

RESULT 14  
US-09-346-237-2  
; Sequence 2, Application US/09346237A  
; Patent No. 6265197  
; GENERAL INFORMATION:  
; APPLICANT: Bisgard-Frantzen, Henrik  
; APPLICANT: Svendsen, Allan  
; TITLE OF INVENTION: Starch Debranching Enzymes  
; FILE REFERENCE: 5629.200-US  
; CURRENT APPLICATION NUMBER: US/09/346,237A  
; CURRENT FILING DATE: 1999-07-01  
; EARLIER APPLICATION NUMBER: PA 1998 00868  
; EARLIER FILING DATE: 1998-07-02  
; EARLIER APPLICATION NUMBER: 60/094,353  
; EARLIER FILING DATE: 1998-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 915  
; TYPE: PRT  
; ORGANISM: Bacillus deramificans  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)..(915)  
; OTHER INFORMATION: Pullulanase  
US-09-346-237-2

Query Match 3.6%; Score 91.5; DB 4; Length 915;  
Best Local Similarity 23.0%; Pred. No. 1.7;  
Matches 79; Conservative 38; Mismatches 127; Indels 99; Gaps 19;

QY 2 DRRPLPVDRAAGLKE--KTLILLDVSTKNP--VRTVNENFLSLQDPSIIHDGWLDFLSS 57  
Db 424 DVRFSDIPNSGMKNKGKYLALTEKGTGKPDNVKT-----GID--SL 463  
QY 58 KRLVTLARGLSPAFLRFGGKRTDFLOFONLRNPAKSRGGPGDYVYLYKNYEDDIIVRSDVAL 117  
Db 464 KQLGTHVQLMPVFNASNSVDETD-----PTQDNWGYDP-----RVN--DVPEGQYAT 508  
QY 118 DKQCKGKTAQHPDVMLEQREKAAQMHVLKLEQFSNTYSNLITPEPN--NYRTMH--GRAV 175  
Db 509 NANGNARIKEFKENKVLHREHIGVNMVYVNHTEFATQISDFDKIVPEYVYRTDAGNYT 568  
QY 176 NGSQGLKNDYIQKLSLLQPIRIYSRASLYSGNIGRPNKVNIALLDGFMKAGSVTDVAT--W 234  
Db 569 NGSGTGNEIAERPMVQKF-----IIDSLLKYW 595  
QY 235 QHCY-IDGRVVKVMDFLKRLLDLSQDQIRKIQKVVNTYTPGKKIWLKLEGVVTTTSAGGTNN 293  
Db 596 VNEYHIDGFRDLMALLGK---DTMS-----KAASELHAINPG--IALYGEPTW--GGTSA 644  
QY 294 LSDSYAAGFLWNLTLGMLANOGIDVIRHFFDHGYNHLVDQN 336  
Db 645 LPDDQ-----LLTKG--AQKMGVAV-----FNDNLRNALDGN 675

RESULT 15  
US-09-514-599-4  
; Sequence 4, Application US/09514599



; Patent No. 6350599  
; GENERAL INFORMATION:  
; APPLICANT: Svendsen, Allan  
; TITLE OF INVENTION: Pullulanase Variants And Methods For  
; TITLE OF INVENTION: Preparing Such Variants With Predetermined Properties  
; FILE REFERENCE: 6072.200-US  
; CURRENT APPLICATION NUMBER: US/09/514,599  
; PRIOR FILING DATE: 2000-02-28  
; PRIOR APPLICATION NUMBER: PA 2000 00045  
; PRIOR FILING DATE: 2000-01-12  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Bacillus deramificans  
US-09-514-599-4

Query Match 3.6%; Score 91.5; DB 4; Length 928;  
Best Local Similarity 23.0%; Pred. No. 1.8;  
Matches 79; Conservative 38; Mismatches 127; Indels 99; Gaps 19;  
  
Qy 2 DRRLPVDRAGLAE--KTLILLDVSTKNP--VRTVNENFLSLQLDPSIIHDGWLDFLSS 57  
Db 437 DVRFSDPNPCKMKNKGYLALTEKGTGKPDNVKT-----GID--SL 476  
  
Qy 58 KRLVTLARGLSPAFRLFGGKRTDFLQONLRNPAKSRGGPGDPYVYKKNYEDDIVRSDVAL 117  
Db 477 KQLGITHVQLMPVFNASNSVDETD-----PTQDNNGYDP---RNY--DVPEGQYAT 521  
  
Qy 118 DKQKGCKIAQHPDVMLELQREKAQMHLVLLKEQFSNTYSNLIITEPN-NYRTMH-GRAY 175  
Db 522 NANGNARIKEFKEMVLSLHREHIGVNMDDVYVNHHTFATQISDFDKIVPEYYRTDDAGNYT 581  
  
Qy 176 NGSQIGKDYIQLSLQPIRIYSRASLYGPNIGRPNKVNIALLDGFMKVGASTVDVAVT-W 234  
Db 582 NGSGTGNEIAAERPVMQKF-----IDSLSKYW 608  
  
Qy 235 QHCY-IDGRVYKVMDFLTKRLTLDSQIRKIQKVNTYTPGKKIWLKGVVTTTSAGGTNN 293  
Db 609 VNEYHIDGFRFDLMALLCK---DTMS-----KAASELHAINFG--IALYGEPTW--GGTSA 657  
  
Qy 294 LSDSYAGFLWNLTLGMLANQIDVVIHRSFFDHGYNHLVDON 336  
Db 658 LPDDQ-----LLATG--AQKGMGVAV-----FNDNLRNALDGN 688

Search completed: July 30, 2002, 08:15:04  
Job time: 286 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:14:02 ; Search time 141.47 Seconds  
(without alignments)  
317.982 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_130\_534  
Perfect score: 2129  
Sequence: 1 NPAKSRGGPGDYLLKNYED.....PVTMGFFVKNVNALACRYR 405

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2129	100.0	534	22	AAB85216	Heparanase-like pr
2	2122	99.7	492	22	AAB84664	Amino acid sequenc
3	2090	98.2	592	22	AAU07424	Human heparanase-1
4	2090	98.2	592	22	AAB85215	Heparanase-like pr
5	2090	98.2	592	22	AAU97632	Human heparanase,
6	2080	97.7	592	22	AAB81062	Human heparanase-2
7	1829	85.9	538	22	AAU97633	Human heparanase,
8	1822	85.6	439	22	AAU07423	Human heparanase-1
9	1822	85.6	480	22	AAU07418	Novel human extrac
10	1822	85.6	480	22	AAB85217	Heparanase-like pr
11	1822	85.6	480	22	AAU97634	Human heparanase,

12	1112	52.2	214	22	AAM99905	Human excretory re
13	1112	52.2	214	22	AAAM3704	Human bladder anti
14	805	37.8	380	20	AAU17085	Rat heparanase enz
15	799.5	37.6	543	20	AAU02345	A human heparanase
16	799.5	37.6	543	21	AAB08849	Amino acid sequenc
17	799.5	37.6	543	21	AAU57590	Human heparanase.
18	799.5	37.6	543	21	AAU52990	Human heparanase p
19	799.5	37.6	543	22	AAU97635	Human heparanase p
20	799.5	37.6	592	20	AAU02346	A human heparanase
21	799.5	37.6	592	21	AAB08850	Amino acid sequenc
22	796.5	37.4	530	20	AAU34173	Human pre-prohepar
23	796.5	37.4	543	22	AAB88361	Human membrane or
24	795.5	37.4	543	20	AAU17082	Human heparanase e
25	795.5	37.4	543	20	AAB86206	Human heparanase i
26	795.5	37.4	588	20	AAU30124	A human protein wi
27	790	37.1	535	21	AAB08851	A murine heparanas
28	779	36.6	380	20	AAU17084	Mouse heparanase e
29	778	36.5	156	22	AAG65963	Human heparanase-1
30	752.5	35.3	532	20	AAU17083	Seq ID No: 15 of W
31	658	30.9	262	22	AAU24147	Human EST encoded
32	410	19.3	488	22	AAB31469	Amino acid sequenc
33	388	18.2	488	22	AAB31470	Amino acid sequenc
34	383	18.0	488	22	AAB31472	Amino acid sequenc
35	375	17.6	488	22	AAB31471	Amino acid sequenc
36	278	13.1	137	22	AAG65961	Human heparanase-1
37	278	13.1	159	22	AAG65964	Human heparanase-1
38	270	12.7	112	22	AAU07425	Human heparanase-1
39	247	11.6	256	21	AAG13479	Arabidopsis thalia
40	203	9.5	118	22	AAG65962	Mouse heparanase-1
41	162	7.6	115	22	AAB85218	Mouse heparanase-1
42	155	7.3	935	22	ABB69219	Drosophila melanog
43	144	6.8	38	20	AAU34186	Human pre-prohepar
44	127	6.0	24	22	AAB85220	Heparanase-like pr
45	127	6.0	25	22	AAB85221	Heparanase-like pr

ALIGNMENTS

RESULT 1  
AAB85216  
ID AAB85216 standard; Protein; 534 AA.  
XX AC AAB85216;  
XX AC  
XX 07-SEP-2001 (first entry)  
XX DE Heparanase-like protein Hpa2 splice variant #2.  
XX DE  
XX KW Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;  
XX KW cytosolic; neuroprotective; cerebroprotective; immunosuppressive;  
XX KW antipsoriatic; nontropic; antiinflammatory; antiarthritic; antiasthmatic;  
XX KW antidiabetic; antiarteriosclerotic; vulneryary.  
XX OS Homo sapiens.  
XX XX  
XX WO200146392-A2.  
XX PD 28-JUN-2001.  
XX XX  
XX PF 21-DEC-2000; 2000WO-GB04963.  
XX PR 22-DEC-1999; 99GB-0030392.  
XX PR 07-APR-2000; 2000GB-0008713.  
XX PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
XX XX  
XX PI McKenzie EA, Stamps AC, Terrett JA, Tyson KL;  
XX WPI; 2001-418056/44.  
XX DR N-PSDB; AAH22672.  
XX XX  
XX PT Novel homologs of heparanase, present in three splice variants, useful

PT	for identifying agents that modulate heparanase, useful in the treatment and/or prophylaxis of abnormal levels of heparanase	
XX	Claim 1; Fig 2; 97pp; English.	
XX	The invention provides a homologue to heparanase which is present in three splice variants. The heparanase homologue polypeptide is useful in the treatment of a human or non-human animal or for use in diagnosis.	
CC	Vectors comprising the heparanase homologue polynucleotides are useful in the transformation or transfection of a prokaryotic or eukaryotic host.	
CC	The modulators of the polypeptide are useful in the manufacture of a medicament for the treatment and/or prophylaxis of a condition/disease associated with abnormal levels of the heparanase homologue, including cancer, central nervous system (CNS) and neurodegenerative diseases,	
CC	cardiovascular diseases such as restenosis following angioplasty and atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,	
CC	allergies, inflammatory diseases, arthritis, vascular restenosis, tumour growth and progression, asthma, Alzheimer's disease, diabetic retinopathy, wound healing and inflammation. The polypeptide is also useful in diagnosis and research. The present sequence represents the amino acid sequence of the mid-sized splice variant of the heparanase-like protein Hpa2 of the invention.	
XX	Sequence 534 AA;	
SQ		
PT	Query Match 100.0%; Score 2129; DB 22; Length 534;	
XX	Best Local Similarity 100.0%; Pred. No. 5.5e-220;	
PS	Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
XX		
CC	1 NPAKSRGGPGDYLLKNEYEDDIVRSVDALDKQCKIAQHDPVMLELOREKAAQHMLVLL 60	
DB	130 npaksrggpgdyllkneyeddivrsvaldkqckiaqhpdmvmlgrekaqmhlvl 189	
QY	61 KEQFSNTYSNLIITPEPNYRTMHGRAVNSQLGKDYIQLKSLLOPIRYSRASLYGPNIG 120	
DB	190 keqfsntysnllitepnnyrtmhgravnsgqlgkdyiklsllqpirysrasygnlg 249	
QY	121 RPRKNVIALDGMKVGSTVDVAVTWQHCYIDGRVVKVMDFLKTRLLDLSQIRKIQRV 180	
DB	250 rprknvialldgmkvagstvdatvqhcyidgrvvkvmfdiktrlldtldsqirkiqkv 309	
QY	181 VNTYTPGCKKWLKLGVTTSAGTNNLSDSYAAGFLWMLNTGLMANGIDVVRHRSFFDHG 240	
DB	310 vntytpgkklwlegvvttsaggttnnlsdsyaagflwntglmlangidvvrhrsffdhg 369	
QY	241 YNHLVDQNFPLDPDYWLSLLYKRLIGPKVLAVHVGALQRPGRVIRDKLRIYAHCTNH 300	
DB	370 ynhlvdqnfpldpdywlslllykrligpkvlavhvaglqrprgvrirdklriyahctnh 429	
QY	301 HHNNYVRGSSITLFTINLHRSRKKIKLAGTLRDKLVHGYLLQPYQGBGLSKSVQLNGQPL 360	
DB	430 hnnnyvrsgstlftinlhrsrrkkiklagtlrdklvhyllqpyqgqglksksvqlngqpl 489	
QY	361 VMVDDGTLPELKPRLRAGRTLPIPPVTMGFFVVKVNNVALACRYR 405	
DB	490 vmvddgtlpeclkprlragrtlppvtmgffvkvnnvalacryr 534	
RESULT 2		
AAB84664		
ID	AAB84664 standard; Protein; 492 AA.	
XX	AAB84664;	
AC		
XX	05-SEP-2001 (first entry)	
DT		
XX	Amino acid sequence of human heparanase-like polypeptide.	
DE		
XX	Human; heparanase-like polypeptide; gene therapy; cancer; angiogenesis;	
KW	trauma; autoimmune disease; skin disease; cardiovascular disease;	
KW	nervous system disease; inflammation; arthritis; genitalia;	
KW	male fertility; erectile dysfunction.	

XX	Homo sapiens.	
OS		
XX	Key Location/Qualifiers	
FH	Misc-difference 407	
FT	/note= "unspecified residue encoded by KCA"	
FT		
XX	WO200148161-A2.	
PN		
XX	05-JUL-2001.	
PD		
XX	18-DEC-2000; 2000WO-EP12909.	
PF		
XX	23-DEC-1999; 99EP-0125831.	
PR		
XX	(SCHD ) SCHERING AG.	
PA		
XX	Siemeister G, Weiss B;	
PI		
XX	WPI: 2001-418259/44.	
DR	N-PSDB; AAH28347.	
XX	Human Heparanase-like polynucleotide encoding polypeptides useful for modulating expression of the polypeptide and for treating cancer, cancer metastasis, aberrant angiogenesis by gene therapy technique -	
PT		
PT	Claim 9; Page 30; 30pp; English.	
XX		
PS	The present sequence represents a human heparanase-like polypeptide.	
CC	Heparanase-like polynucleotides are useful as a source of probes, like primers and antisense molecules, and in gene therapy. Heparanase-like polynucleotides and polypeptides are useful for treating several disorders e.g., cancer, cancer metastasis. The oligonucleotides are also useful as diagnostic markers for the diagnosis of disorder such as cancer, cancer metastasis and aberrant angiogenesis. They may also act as diagnostic markers for diagnosis of disorder such as cancer, cancer metastasis and aberrant angiogenesis. The heparanase polypeptides and polynucleotides are also useful for treating trauma, autoimmune diseases, skin diseases, cardiovascular diseases, nervous system diseases, and inflammation including arthritis. Since the polynucleotide is preferentially expressed in male genitalia, modulation of its expression and/or activity may be used for medical intervention in male genitalia function that is male fertility control, erectile dysfunction.	
CC	Sequence 492 AA;	
XX		
SQ		
Query Match	99.7%; Score 2122; DB 22; Length 492;	
Best Local Similarity	99.5%; Pred. No. 2.7e-219;	
Matches	403; Conservative 1; Mismatches 1; Indels 0; Gaps 0;	
QY	1 NPAKSRGGPGDYLLKNEYEDDIVRSVDALDKQCKIAQHDPVMLELOREKAAQHMLVLL 60	
DB	88 npaksrggpgdyllkneyeddivrsvaldkqckiaqhpdmvmlgrekaqmhlvl 147	
QY	61 KEQFSNTYSNLIITPEPNYRTMHGRAVNSQLGKDYIQLKSLLOPIRYSRASLYGPNIG 120	
DB	148 keqfsntysnllitepnnyrtmhgravnsgqlgkdyiklsllqpirysrasygnlg 207	
QY	121 RPRKNVIALDGMKVGSTVDVAVTWQHCYIDGRVVKVMDFLKTRLLDLSQIRKIQRV 180	
DB	208 rprknvialldgmkvagstvdatvqhcyidgrvvkvmfdiktrlldtldsqirkiqkv 267	
QY	181 VNTYTPGCKKWLKLGVTTSAGTNNLSDSYAAGFLWMLNTGLMANGIDVVRHRSFFDHG 240	
DB	268 vntytpgkklwlegvvttsaggttnnlsdsyaagflwntglmlangidvvrhrsffdhg 327	
QY	241 YNHLVDQNFPLDPDYWLSLLYKRLIGPKVLAVHVGALQRPGRVIRDKLRIYAHCTNH 300	
DB	328 ynhlvdqnfpldpdywlslllykrligpkvlavhvaglqrprgvrirdklriyahctnh 387	
QY	301 HHNNYVRGSSITLFTINLHRSRKKIKLAGTLRDKLVHGYLLQPYQGBGLSKSVQLNGQPL 360	

Db 388 hnmvyrsgitlfiinlhrxrkkiklagtldklvghqyllqpygqeglkksvqlngqpl 447

Qy 361 VMVDDGTLPELKPRPLRAGRTLVIPPTVMGFVVKVNNALACRYR 405  
|||||

Db 448 vmvddgtlpeikprlragrtlvipptvmgfyvkvnnalacryr 492

RESULT 3  
AAU07424  
ID AAU07424 standard; Protein; 592 AA.  
XX AC AAU07424;  
XX DT 18-DEC-2001 (first entry)  
XX DE Human heparanase-like protein splice variant #1.  
XX KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
KW antiproliferative; cardiac; vasotropic; cerebroprotective; neurotropic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;  
KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;  
KW wound healing; food additive; heparanase.  
XX OS Homo sapiens.  
XX WO200179253-A1.  
XX PN 25-OCT-2001.  
XX PD 11-APR-2001; 2001WO-US11643.  
XX PF 18-APR-2000; 2000US-198123P.  
XX PR (HUMA-) HUMAN GENOME SCI INC.  
XX PA Flscella M, Shi Y, Ebner R, Ruben SM;  
XX PI WPI; 2001-611720/70.  
XX DR N-PSDB; AAS13848.  
XX DR New nucleic acids encoding extracellular matrix polypeptides, for  
PT diagnosing, treating, preventing or ameliorating human disorders and  
PT disease, such as, autoimmune, hyperproliferative or cardiovascular  
PT disorders -  
XX PS Disclosure; Page 14; 308pp; English.

CC The invention relates to novel isolated polynucleotides (I) encoding  
CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by  
CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.  
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility to  
CC a pathological condition. The antibodies to the polypeptides can also be  
CC used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities. The present sequence represents the amino acid sequence  
CC of human heparanase-like protein, splice variant #1.

SQ Sequence 592 AA;

Query Match 98.2%; Score 2090; DB 22; Length 592;  
Best Local Similarity 87.5%; Pred. No. 1e-215; 0; Indels 58; Gaps 1;  
Matches 405; Conservative 0; Mismatches 0;

QY 1 NPAKSRGGPGDYLYLKNYEDDIVRSDVALDKQKGIQAQHPDVMLEQREKAQMHVL 60  
|||

Db 130 npakserggpgdylylknayeddivrsdvaldkqkgkiaghpvmleqrekaqmhvl 189  
|||

QY 61 KEQFSNTYSNLILT----- 74  
|||

Db 190 keqfsntysnllltarsldklynfadcsghlalfalnrrnnpnswssalsllkysa 249  
|||

QY 75 -----EPNNYRTMHGRAVNGSOLGKDYIQLKSLLOPIRYSRASLYGPNIGRP 122  
|||

Db 250 skkyniswelgnepnmyrtmhgravngslgkdyqlkllqpiriyraslygpnigrp 309  
|||

QY 123 RKNVIALLDGFMKVGASTVDVMTQHCYIDGRVVKVMDFLKTRLLDLSQIRKIQKVVN 182  
|||

Db 310 rknavialldgfmkvagstvdvtwqhcylidgrvkvkvmdfiktrlldtsdqirkikvvn 369  
|||

QY 183 TYTPGKKIWLGVVTTSGAGTNNLSDSYAAGFLWLNTLGLANOGIDVIRHSFFDHGYN 242  
|||

Db 370 tytpgkkiwlegvvttsaggttnlsdyaagflwntlgmlanqgidvvrhsffdhgyn 429  
|||

QY 243 HLVDQNFNLPDYWLISLLYKRLIGPKVLAVHVGALQKRPGRVIRDKLRIYAHCTNHNH 302  
|||

Db 430 hlvdqnfnpdywlislykrligpkvlavhvgalqkrgpgrvirkkiriyahecnhnh 489  
|||

QY 303 HNYVRGSITLFIINLHRSRKKIKLAGTRDKLVHQYLLQPYGOEGLKSKSVQLNGQPLVM 362  
|||

Db 490 hnyvrgsitlfiinlhrrs-kkiklagtldklvghqyllqpygqeglkksvqlngqplvm 549  
|||

QY 363 VDDGTLPELKPRPLRAGRTLVIPPTVMGFVVKVNNALACRYR 405  
|||

Db 550 vddgtlpeikprlragrtlvipptvmgfyvkvnnalacryr 592  
|||

RESULT 4  
AAB85215  
ID AAB85215 standard; Protein; 592 AA.  
XX AC AAB85215;  
XX DT 07-SEP-2001 (first entry)  
XX DE Heparanase-like protein Hpa2 splice variant #1.  
XX KW Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;  
KW cytostatic; neuroprotective; cerebroprotective; immunosuppressive;  
KW antipsoriatic; neutropic; antiinflammatory; antiarthritic; antiasthmatic;  
KW antidiabetic; antiarteriosclerotic; vulnerary.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
FT Misc-difference 237 /label= unknown  
FT /note= "encoded by ANC"  
XX WO200146392-A2.  
XX PD 28-JUN-2001.  
XX PF 21-DEC-2000; 2000WO-GB04963.  
XX PR 22-DEC-1999; 99GB-0030392.  
XX PR 07-APR-2000; 2000GB-0008713.  
XX PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
XX

PI McKenzie EA, Stamps AC, Terrett JA, Tyson KL;  
XX WPI; 2001-418056/44.  
DR N-PSDB; AAH22671.

XX Novel homologs of heparanase, present in three splice variants, useful  
PT for identifying agents that modulate heparanase, useful in the  
PT treatment and/or prophylaxis of abnormal levels of heparanase -  
XX  
PS  
PS Claim 1; Fig 1; 97pp; English.

CC The invention provides a homologue to heparanase which is present in  
CC three splice variants. The heparanase homologue polypeptide is useful in  
CC the treatment of a human or non-human animal or for use in diagnosis.  
CC Vectors comprising the heparanase homologue polynucleotides are useful in  
CC the transformation or transfection of a prokaryotic or eukaryotic host.  
CC The modulators of the polypeptide are useful in the manufacture of a  
CC medicament for the treatment and/or prophylaxis of a condition/disease  
CC associated with abnormal levels of the heparanase homologue, including  
CC cancer, central nervous system (CNS) and neurodegenerative diseases,  
CC cardiovascular diseases such as restenosis following angioplasty and  
CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,  
CC allografts, inflammatory diseases, arthritis, vascular restenosis,  
CC tumour growth and progression, asthma, Alzheimer's disease, diabetic  
CC retinopathy, wound healing and inflammation. The polypeptide is also  
CC useful in diagnosis and research. The present sequence represents the  
CC amino acid sequence of the largest splice variant of the heparanase-like  
CC protein Hpa2 of the invention.

XX Sequence 592 AA;

Query Match 98.2%; Score 2090; DB 22; Length 592;  
Best Local Similarity 87.5%; Pred. No. 1e-215;  
Matches 405; Conservative 0; Mismatches 0; Indels 58; Gaps 1;

QY 1 NPAKSRGGPGDYLLKNYEDDIVRSDVALDKQKCKIAQHPDVMLELQREKAAQHLVLL 60  
DB 130 npaksrggpgdyllknyeddivrsdvaldkqkckiaqhpdvmlqlgrekaqmhlvl 189  
QY 61 KEQFSNTYSNLIIT----- 74  
DB 190 keqfsntysnliitarsldklynfadcsglhlfalnrrnppnswxssalsllkysa 249  
QY 75 -----EPNNYRTMHGRAVNGSOLGKDYIQLKSLLOPIRYSRASLYGPNIGRP 122  
DB 250 skkyniswelgnepnyrtmhgravnsglqkdyiqklsllqipirysraslygpnigrp 309  
QY 123 RKNVIALLDGFMKVGASTVDVATWQHICYIDGRVVKVMDFLKTRLLDLSQIRKIQKVVN 182  
DB 310 rknvialldgfmkvagstvdatwqhcyidgrvvkvmdflktrlldtldsqirkikqvv 369  
QY 183 TYTPGKKIWLGVVTTTSAGTNNLSDSYAGFLWLNTLGLMANOGIDVVIHRSFFDHGYN 242  
DB 370 tytpgkklwlegvvtttsagtnnlsdsyaagflwntlgmlanogidvvihrsfddhgn 429  
QY 243 HLVDQNFNPLPDYWLSLYKRLIGPKVLAVHVAGLQKPRPGVIRDKLRIYAHCTNNHN 302  
DB 430 hlvdqnfnpdpdywlslykrligpkvlavhvaglrqprgvrirdklriyahctnnhn 489  
QY 303 HNYVRGSITLFIINLHRSRKKIAGTFLRDLKLVHQYLLQPYGGEGLSKSVQINGPLVM 362  
DB 490 hnyvrgsitlfiinlhrrskkiklagtflrldklvqhyyllqpyggeglsksvqngqplvm 549  
QY 363 VDDGTLPELKPRLPRLAGRTLVIPVTMGFFVKNVNALACRYR 405  
DB 550 vddgtlpeelprrlprlragrtlvipvtmgffvknvnalacryr 592

RESULT 5  
AAY97632  
ID AAY97632 standard; Protein; 592 AA.  
XX

AC AAY97632;  
XX 20-APR-2001 (first entry)  
DT  
XX Human heparanase, hnhpl, protein sequence.  
DE  
XX Heparanase: hnhpl; wound healing; angiogenesis; restenosis; Scrape;  
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;  
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;  
KW gene therapy; human.  
XX  
XX Homo sapiens.  
OS  
XX WO200100643-A2.  
PN  
XX 04-JAN-2001.  
PD  
XX 19-JUN-2000; 200QWO-IL00358.  
PF  
XX 25-JUN-1999; 99US-0140801.  
PR (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
XX  
XX Pecker I, Michal I, Itzhaki H;  
PI  
XX WPI; 2001-137930/14.  
DR N-PSDB; AAA91097.  
DR  
XX New polynucleotides and polypeptides that are distantly homologous to  
PT heparanase, useful in wound healing, as well as in gene therapy  
PT protocols for angiogenesis, restenosis, atherosclerosis, or  
PT inflammation -  
PT  
XX Claim 10; Fig 1; 67pp; English.

PS This sequence represents a heparanase of the invention.  
XX The heparanase DNA and protein sequences are useful in wound healing,  
CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary  
CC diseases, neurodegenerative diseases (such as Scrape, Alzheimer's  
CC disease, and Creutzfeldt-Jakob disease) or viral infections. The  
CC heparanase coding sequence is particularly useful in gene therapy.

XX Sequence 592 AA;

Query Match 98.2%; Score 2090; DB 22; Length 592;  
Best Local Similarity 87.5%; Pred. No. 1e-215;  
Matches 405; Conservative 0; Mismatches 0; Indels 58; Gaps 1;

QY 1 NPAKSRGGPGDYLLKNYEDDIVRSDVALDKQKCKIAQHPDVMLELQREKAAQHLVLL 60  
DB 130 npaksrggpgdyllknyeddivrsdvaldkqkckiaqhpdvmlqlgrekaqmhlvl 189  
QY 61 KEQFSNTYSNLIIT----- 74  
DB 190 keqfsntysnliitarsldklynfadcsglhlfalnrrnppnswxssalsllkysa 249  
QY 75 -----EPNNYRTMHGRAVNGSOLGKDYIQLKSLLOPIRYSRASLYGPNIGRP 122  
DB 250 skkyniswelgnepnyrtmhgravnsglqkdyiqklsllqipirysraslygpnigrp 309  
QY 123 RKNVIALLDGFMKVGASTVDVATWQHICYIDGRVVKVMDFLKTRLLDLSQIRKIQKVVN 182  
DB 310 rknvialldgfmkvagstvdatwqhcyidgrvvkvmdflktrlldtldsqirkikqvv 369  
QY 183 TYTPGKKIWLGVVTTTSAGTNNLSDSYAGFLWLNTLGLMANOGIDVVIHRSFFDHGYN 242  
DB 370 tytpgkklwlegvvtttsagtnnlsdsyaagflwntlgmlanogidvvihrsfddhgn 429  
QY 243 HLVDQNFNPLPDYWLSLYKRLIGPKVLAVHVAGLQKPRPGVIRDKLRIYAHCTNNHN 302  
DB 430 hlvdqnfnpdpdywlslykrligpkvlavhvaglrqprgvrirdklriyahctnnhn 489

Qy 303 HNYVRSITLFIINLHRSRKKIKIAGTLRDKLVHQLVLPQYQGEGLSKSVOLNGQPLVM 362  
Db 490 hnyvrgsitfiinlhrrskkiklagtlrdklvbgyllyqpygqeglsksvqlngqplvm 549  
Qy 363 VDDGTLPELKPRLRAGRTLVIPTVMGFFVVKVNNALACRYR 405  
Db 550 vddgtlpeikprlragrtlviptvmgffvfvkvnalacryr 592

RESULT 6  
AAB81062  
ID AAB81062 standard; Protein: 592 AA.  
XX AAB81062;  
XX  
DT 20-JUN-2001 (first entry)  
XX  
DE Human Heparanase-2 amino acid sequence.  
XX  
KW Heparanase 2; human; endoglucuronidase; heparan sulphate; metastasis;  
KW neocangiogenesis; vaccine; autoimmune disorder; blood coagulation;  
KW cancer; diabetes; ischaemia; sepsis; stroke; cardiovascular; thrombosis.  
XX  
OS Homo sapiens.

XX Key Location/Qualifiers  
FH Region 156..169  
FT /label= Immunogenic\_epitope  
FT Region 249..262  
FT /label= Immunogenic\_epitope  
FT Region 505..518  
FT /label= Immunogenic\_epitope

XX WO200121814-A1.  
XX 29-MAR-2001.  
XX  
XX 11-SEP-2000; 2000WO-EP08837.  
XX  
XX 23-SEP-1999; 99EP-0118805.  
PR 07-JUL-2000; 2000EP-0114649.  
XX  
XX (MERE ) MERCK PATENT GMBH.  
XX  
XX Duecker K, Sirrenberg C;  
XX  
XX WPI: 2001-308089/32.  
DR N-PSDB: AAF86101.  
XX  
XX New heparanase-2 polypeptide useful in diagnosing (the susceptibility  
PT of a subject to) and as vaccines against e.g. autoimmune disorders, or  
PT cardiovascular disease, cancer, diabetes, ischemia, sepsis, stroke, or  
PT thrombosis  
XX  
XX Claim 1; Page 42-43; 46pp; English.

XX This invention relates to a human heparanase-2 protein and the cDNA  
XX sequence encoding it. Heparanase-2 is a member of the endoglucuronidase  
CC family of polypeptides and it degrades heparan sulphate proteoglycans and  
CC HSPGs (ubiquitous macromolecules of cell surfaces, basement membranes and  
CC the extracellular matrix). HSPGs support the vascular endothelium and  
CC stabilise the structure of the capillary wall. Heparanases may be  
CC associated with neocangiogenesis and metastasis related to malignant  
CC tumour formation. Heparanase-2 polynucleotides and proteins are useful as  
CC vaccines for inducing an immunological response against autoimmune  
CC disorders, blood coagulation disorders, cancer, diabetes, ischaemia,  
CC sepsis, stroke, cardiovascular diseases, or thrombosis, as well as in  
CC diagnosing (the susceptibility of a subject to) these diseases.  
CC Heparanase-2 fragments may be used as immunogens to produce antibodies  
CC immunospecific to the polypeptides, and to identify membrane bound  
CC soluble receptors, agonists or antagonists that compete with the binding  
CC of the polypeptide to the receptors. An antibody specific for  
CC heparanase-2 can be used in the diagnosis of the above diseases and in

CC isolating or identifying clones expressing heparanase-2. The present  
CC sequence represents heparanase-2. Three regions of heparanase-2 with high  
CC immunogenicity (immunogenic epitopes) can be used to raise antibodies  
CC against heparanase-2.  
XX  
SQ Sequence 592 AA;

Query Match 97.7%; Score 2080; DB 22; Length 592;  
Best Local Similarity 87.0%; Pred. No. 1.2e-214;  
Matches 403; Conservative 1; Mismatches 1; Indels 58; Gaps 1;  
Qy 1 NPAKSRGGPGDYLLKNEYEDDIVRSDVALDKOKGCKIAHQPDVMLELQREKAAQHLLVLL 60  
Db 130 npaksrvgpgpdyllknyeddivrsdvaldkqkgckiaqhpdmvlvlgrekaaqmhlvl 189  
Qy 61 KEQFSNTYSNLILT----- 74  
Db 190 keqfsntysnlltarsldklynfadcsglhlifalnrlrnpnswssalsilkysa 249  
Qy 75 -----EPNNYRTMHGRAVNGSOLGKDYIQLKSLIQPIRIYSRASLYGPNIGRP 122  
Db 250 skkyniswelgnepnnyrtmhgravnsglqgkdyqlksllqpiriysraslygpnigrp 309  
Qy 123 RKNVIALLDGFMKVAGSTVDVAVTWQHCYIDGRVVKVMDFLKTRLLDTLSDQIRKIQKVVN 182  
Db 310 rknvialldgfmkvagstvdavtwqhcyidgrvvkvmdfkltrlldtsdqirkikgvvn 369  
Qy 183 TYTPGKKIWLQGVVTTTSAGTNNLSDSYAAGFLWLTGLMLANQIGDIVVIRISFDFHGVN 242  
Db 370 tytpgkklwlegvvttsaggtnnlsdsyaagflwltgmlanqigdivvirsfdhgvn 429  
Qy 243 HLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVAGLQRPGRVIRDKLRIYAHCTNHHN 302  
Db 430 hlvdnfnplpdywlslllykrligpkvlavhvaglqrprgvirdklriyahctnhhn 489  
Qy 303 HNYVRSITLFIINLHRSRKKIKLAGTLRDKLVHQLVLPQYQGEGLSKSVOLNGQPLVM 362  
Db 490 hnyvrgsitfiinlhrrskkiklagtlrdklvbgyllyqpygqeglsksvqlngqplvm 549  
Qy 363 VDDGTLPELKPRLRAGRTLVIPTVMGFFVVKVNNALACRYR 405  
Db 550 vddgtlpeikprlragrtlviptvmgffvfvkvnalacryr 592

RESULT 7  
AAY97633  
ID AAY97633 standard; Protein: 538 AA.  
XX  
XX AAY97633;  
XX  
XX 20-APR-2001 (first entry)  
XX  
XX Human heparanase, hnhpl pn9 form, protein sequence.  
XX  
XX Heparanase; hnhpl; wound healing; angiogenesis; restenosis; Scrape;  
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;  
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;  
XX gene therapy; human.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FT Misc-difference 305 /note= "encoded by GAC"  
XX  
XX WO200100643-A2.  
XX  
XX 04-JAN-2001.  
XX  
XX 19-JUN-2000; 2000WO-IL00358.  
XX  
XX 25-JUN-1999; 99US-0140801.

XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX PI Pecker I, Michal I, Itzhaki H;

XX XX WPI; 2001-137930/14.

DR DR N-PSDB; AAA91098.

XX XX New polynucleotides and polypeptides that are distantly homologous to

PT PT heparanase, useful in wound healing, as well as in gene therapy

PT PT protocols for angiogenesis, restenosis, atherosclerosis, or

PT PT inflammation -

XX XX Claim 10; Page 61-62; 67pp; English.

XX XX This sequence represents a heparanase of the invention.

CC CC The heparanase DNA and protein sequences are useful in wound healing,

CC CC anglogenesis, restenosis, atherosclerosis, inflammation, pulmonary

CC CC diseases, neurodegenerative diseases (such as Scrape, Alzheimer's

CC CC disease, and Creutzfeldt-Jakob disease) or viral infections. The

CC CC heparanase coding sequence is particularly useful in gene therapy.

XX XX Sequence 538 AA;

Query Match 85.9%; Score 1829; DB 22; Length 538;

Best Local Similarity 88.3%; Pred. No. 1.2e-187;

Matches 364; Conservative 2; Mismatches. 36; Indels 10; Gaps 4;

QY 1 NPAKSRGGPGPDYLLKNYEDDIVRSDVALDKQCKIAQHPDVMLEQREKA-----AQM 55

DB 130 npaksggpgpdyllknyed--arsldklynfadcs-gihllifalnairnpnswns 186

QY 56 HLVLLEKESNTY--SNLIITPEPNYRTMHGRAVNGSQLGKDYIQLKSLQPIRYSRAS 113

DB 187 alsllkysaskynisweignepnnyrtmhgravgngsqigkdyiqikslqpirlystras 246

QY 114 LYPGNTGRPRKNVIALLDGFMKVGASTVDVATWQHGYIDGRVVKVMDFLKTRLLDTLSQ 173

DB 247 lygpnlgprkpnvialldgfmkvagstvavtwqhgyidgrvkvmdflktrlldtlsa 306

QY 174 IRKIQKVVNTYTPGKKIWLGEVVTTSAGGTNNLSDSYAGFLWNTLGLMANQGDVVR 233

DB 307 irkikqvntypgkklwlegvvttsaggtnnlsdsyagflwntlglmanqgdvvr 366

QY 234 HSEFDHCYNHLVDQNFPLDYLLKRLIGPKVLAVHAGLQKPRGVRIRDKLRI 293

DB 367 hsfidhgyhnlvdqnfpldyllskrligpkvlavhaglqkprgvrirdklri 426

QY 294 YAHCTNHHNHYVRGSTITFLIINLHRSRKKIKLAGTLRDKLVHQLQPYQEGLSKSV 353

DB 427 yahctnhhnhvrgstiflinlhrsrrkkiklagtlrdklvqhyllypygqeglsksv 486

QY 354 QLNGQPLVWDDGTLPELKPRLRAGTLVTPVTWGEFFVVKVNNALACRYR 405

DB 487 qlngqpvmvddgtlpeikprlragnetlvtpvtwgmffvkvnnalacryr 538

RESULT 8

AAU07423

XX ID AAU07423 standard; Protein; 439 AA.

XX AC AAU07423;

XX XX 18-DEC-2001 (first entry)

XX DT Human heparanase-like protein.

XX DE Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

XX KW antiproliferative; cardiant; vasotropic; cerebroprotective; neotropic;

XX KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;

XX KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;

XX KW hyperproliferative disorder; neoplasm; cardiovascular disorder;

KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;

KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;

XX wound healing; food additive; heparanase.

XX OS Homo sapiens.

XX XX WO200179253-A1.

XX XX 25-OCT-2001.

XX XX 11-APR-2001; 2001WO-US11643.

XX XX 18-APR-2000; 2000US-198123P.

XX XX (HUMA-) HUMAN GENOME SCI INC.

XX XX Fiscella M, Shi Y, Ebner R, Ruben SM;

XX XX WPI; 2001-611720/70.

XX XX New nucleic acids encoding extracellular matrix polypeptides, for

PT PT diagnosing, treating, preventing or ameliorating human disorders and

PT PT disease, such as, autoimmune, hyperproliferative or cardiovascular

PT PT disorders -

XX XX Disclosure; Page 13-14; 308pp; English.

XX XX The invention relates to novel isolated polynucleotides (I) encoding

CC CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by

CC CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.

CC CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They

CC CC are also used in diagnosing a pathological condition or susceptibility to

CC CC a pathological condition. The antibodies to the polypeptides can also be

CC CC used in alleviating symptoms associated with the disorders and in

CC CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked

CC CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated

CC CC include autoimmune diseases e.g. rheumatoid arthritis,

CC CC hyperproliferative disorders e.g. neoplasms of the breast or liver,

CC CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders

CC CC e.g. cerebral ischaemia, angioneurosis, nervous system disorders e.g.

CC CC Alzheimer's disease, infections caused by bacteria, viruses and fungi

CC CC and ocular disorders e.g. corneal infection. The polypeptides can also

CC CC be used to aid wound healing and epithelial cell proliferation, to

CC CC prevent skin aging due to sunburn, to maintain organs before

CC CC transplantation, for supporting cell culture of primary tissues, to

CC CC regenerate tissues and in chemotaxis. The polypeptides can also be used

CC CC as a food additive or preservative to increase or decrease storage

CC CC capabilities. The present sequence represents the amino acid sequence

CC CC of human heparanase-like protein.

XX XX Sequence 439 AA;

Query Match 85.6%; Score 1822; DB 22; Length 439;

Best Local Similarity 86.7%; Pred. No. 4.7e-187;

Matches 351; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 NPAKSRGGPGPDYLLKNYEDDIVRSDVALDKQCKIAQHPDVMLEQREKAQMHLVLL 60

DB 89 npaksggpgpdyllknyed----- 108

QY 61 KEQFSNTYSNLIITPEPNYRTMHGRAVNGSQLGKDYIQLKSLQPIRYSRASLYGPNIG 120

DB 109 -----epnnyrtmhgravgngsqigkdyiqikslqpirysrasygpnig 154

QY 121 RPRKNVIALLDGFMKVGASTVDVATWQHGYIDGRVVKVMDFLKTRLLDTLSQIRKIQV 180

DB 155 rprknvialldgfmkvagstvavtwqhgyidgrvkvvmfdlktlrlldtldsqirkiqv 214

QY 181 VNTYTPGKKIWLGEVVTTSAGGTNNLSDSYAGFLWNTLGLMANQGDVIRHSEFDDHG 240

DB 215 vntytpgkklwlegvvttsaggtnnlsdsyagflwntlglmanqgdvivrhsfddhg 274



QY 241 YNHLVDQNFNPLPDYWLISLYKRLIGPKVLAVHVAGLQRPORVIRDKLRIYAHCTNH 300  
Db 275 ynhlvdqnfnpdpdywlslykrligpkvlavhvaglqrpqrvirdklriyahctnh 334  
QY 301 HNNHYVSGSTWFTIINLHRSRKKIKLAGTLRDKLVHVOYLQPYGOEGLKSKSVOLNGOPL 360  
Db 335 hnnhyvrgstlftiinlhrsrrkkiklagtlrdklvhyllqpygqglksksvqlngqpl 394  
QY 361 VMVDDGTLPELKPRPLRAGRTLVIPTVMGFFVVKNNNALACRYR 405  
Db 395 vmvddgtlpeikprplragrtlviptvtmgffvkvknvalacryr 439  
RESULT 9  
AAU07418  
ID AAU07418 standard; Protein: 480 AA.  
XX  
AC AAU07418;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Novel human extracellular matrix (ECM) protein #1.  
XX  
KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
KW antiproliferative; cardiant; vasotropic; cerebroprotective; neurotropic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;  
KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;  
KW wound healing; food additive.  
XX  
OS Homo sapiens.  
XX  
XX WO200179253-A1.  
XX  
XX 25-OCT-2001.  
XX  
XX 11-APR-2001; 2001WO-US11643.  
XX  
XX 18-APR-2000; 2000US-198123p.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Fiscella M, Shi Y, Ebner R, Ruben SM;  
XX  
XX WPI: 2001-611720/70.  
XX  
XX N-PSDB: AAS13843.  
XX  
XX New nucleic acids encoding extracellular matrix polypeptides, for  
XX diagnosing, treating, preventing or ameliorating human disorders and  
XX disease, such as, autoimmune, hyperproliferative or cardiovascular  
XX disorders -  
XX  
XX Claim 1; Page 292-293; 308pp; English.  
XX  
XX The invention relates to novel isolated polynucleotides (I) encoding  
XX extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by  
XX (I) are used to prevent, treat or ameliorate a medical condition in e.g.  
XX humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They  
XX are also used in diagnosing a pathological condition or susceptibility to  
XX a pathological condition. The antibodies to the polypeptides can also be  
XX used in alleviating symptoms associated with the disorders and in  
XX diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
XX immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
XX include autoimmune diseases e.g. rheumatoid arthritis,  
XX hyperproliferative disorders e.g. neoplasms of the breast or liver,  
XX cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
XX e.g. cerebral ischaemia, angiodenesis, nervous system disorders e.g.  
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi  
XX and ocular disorders e.g. corneal infection. The polypeptides can also  
XX be used to aid wound healing and epithelial cell proliferation, to  
XX prevent skin aging due to sunburn, to maintain organs before

CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities. The present sequence represents the amino acid sequence  
CC of novel human extracellular matrix (ECM) protein #1.  
XX  
SQ Sequence 480 AA;  
Query Match 85.6%; Score 1822; DB 22; Length 480;  
Best Local Similarity 86.7%; Pred. No. 5.5e-187;  
Matches 351; Conservative 0; Mismatches 0; Indels 54; Gaps 1;  
QY 1 NPAKSRGGPGPDYLLKNYEDDIVRSQVALDKQKCKIAQHPDVMLELOREKAAQMHVLL 60  
Db 130 npaksggpgpdyllknyed----- 149  
QY 61 KEQFSNTYSNLLITPNNYRTMHGRAVNGSQLGKDYIQLKSLQPIRISYRASLYGPNIG 120  
Db 150 -----epnnyrmtmhgravngsqigkdyiglkslqpirisrasygpnig 195  
QY 121 RPRKNVIALLDGFMKVGAGSTVDVATWQHCYIDGRVVKVMDFLKTLRLDTLSQIRKIOKV 180  
Db 196 rprknvialldgfmkvagstvdavtcqhcyidgrvvkvmdflktrlldtlsdqirkikqv 255  
QY 181 VNTYTPGKKIWLEGYVVTTSAGTNNLSYAAAGFLWLNTLGLANQGDVIVIRHSFFDGH 240  
Db 256 vntytpgkkiwegvvttsagtnnlsdyaagflwntlgmlanqgdvivrhsffdhg 315  
QY 241 YNHLVDQNFNPLPDYWLISLYKRLIGPKVLAVHVAGLQRPGRVIRDKLRIYAHCTNH 300  
Db 316 ynhlvdqnfnpdpdywlslykrligpkvlavhvaglqrpgrvirdklriyahctnh 375  
QY 301 HNNHYVSGSTWFTIINLHRSRKKIKLAGTLRDKLVHVOYLQPYGOEGLKSKSVOLNGOPL 360  
Db 376 hnnhyvrgstlftiinlhrsrrkkiklagtlrdklvhyllqpygqglksksvqlngqpl 435  
QY 361 VMVDDGTLPELKPRPLRAGRTLVIPTVMGFFVVKNNNALACRYR 405  
Db 436 vmvddgtlpeikprplragrtlviptvtmgffvkvknvalacryr 480  
RESULT 10  
AAB85217  
ID AAB85217 standard; Protein: 480 AA.  
XX  
AC AAB85217;  
XX  
DT 07-SEP-2001 (first entry)  
XX  
DE Heparanase-like protein Hpa2 splice variant #3.  
XX  
KW Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;  
KW cytostatic; neuroprotective; cerebroprotective; immunosuppressive;  
KW antipsoriatic; neurotropic; antiinflammatory; antiarthritic; antiasthmatic;  
KW antidiabetic; antiarteriosclerotic; vulnery.  
XX  
OS Homo sapiens.  
XX  
XX WO200146392-A2.  
XX  
PD 28-JUN-2001.  
XX  
PF 21-DEC-2000; 2000WO-GB04963.  
XX  
PR 22-DEC-1999; 99GB-0030392.  
PR 07-APR-2000; 2000GB-0008713.  
XX  
PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
XX  
PI McKenzie EA, Stamps AC, Terrett JA, Tyson KL;  
XX  
DR WPI: 2001-418056/44.

DR N-PSDB; RAH22673.

XX Novel homologs of heparanase, present in three splice variants, useful

PT for identifying agents that modulate heparanase, useful in the

PT treatment and/or prophylaxis of abnormal levels of heparanase -

XX

XX Claim 1; Fig 3; 97pp; English.

XX

CC The invention provides a homologue to heparanase which is present in

CC three splice variants. The heparanase homologue polypeptide is useful in

CC the treatment of a human or non-human animal or for use in diagnosis.

CC Vectors comprising the heparanase homologue polynucleotides are useful in

CC the transformation or transfection of a prokaryotic or eukaryotic host.

CC The modulators of the polypeptide are useful in the manufacture of a

CC medicament for the treatment and/or prophylaxis of a condition/disease

CC associated with abnormal levels of the heparanase homologue, including

CC cancer, central nervous system (CNS) and neurodegenerative diseases,

CC cardiovascular diseases such as restenosis following angioplasty and

CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,

CC allografts, inflammatory diseases, arthritis, vascular restenosis,

CC tumour growth and progression, asthma, Alzheimer's disease, diabetic

CC retinopathy, wound healing and inflammation. The polypeptide is also

CC useful in diagnosis and research. The present sequence represents the

CC amino acid sequence of the smallest splice variant of the heparanase-

CC like protein Hpa2 of the invention.

XX

XX Sequence 480 AA;

Query Match' 85.6%; Score 1822; DB 22; Length 480;

Best Local Similarity 86.7%; Pred. No. 5.5e-187;

Matches 351; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 NPAKSRGGGPDYLLKNYEDDIDVRSVALDKQKGCKIAQHDPVMLELQREKAAQMHVLVL 60

DB 130 npaksggpgpdyllknedy----- 149

QY 61 KEQFSNTYSNLITLTPNNYRTMHGRAVNSQLGKDYIOLKSLLOPIRIYSRASLYGPNIG 120

DB 150 -----epnnryrtmhgravnsgqlgkdyiqklsllqpiriyraslygpnig 195

QY 121 RPRKNVIALLDGFMKVAGSTDAVTWQHICYIDGRVVKVMDFLKRLDPTLSQIRKIOKV 180

DB 196 rprknvialldgfmkvagstdvatwqhcyidgrvvkvmdfklrldtldsqirkiqkv 255

QY 181 VNTYTPGKKIWLEGVVTTSAGTNNLSDSYAAGFLWNTLGLMANGQIDVIRHSHFFDHG 240

DB 256 vntytpgkkiwlegvvttsaggtnnlsdsyaagflwntlglmanqgidvirshffdhg 315

QY 241 YNHLVDQNFNPLPDYWSLLYKRLIGPKVLAVHVAGLQRPGRVIRDKLRIYAHCTNH 300

DB 316 ynhlvdqnfnpldywslllykrlgpkvlavhvaglqrpgrvirdklriyahctnh 375

QY 301 HHNNYVSGSTITFIINLHRSRKKIKIAGTLRDKLVHQYLLQPYGQGLKSKSVQLNGQPL 360

DB 376 hnnnyvrgstifilnlhrrskkiklagtlrdklvhqyllqpygqglksksvqlngqpl 435

QY 361 VMVDDGTLPDLPRPLRAGRTLVIPTVTMGFFVVKVNNALACRYR 405

DB 436 vmvddgtlpeklprlragrtlivpptvtmgffvkvnnalacryr 480

RESULT 11

AA97634

ID AAY97634 standard; Protein; 480 AA.

XX

XX AAY97634;

XX AC

XX DT

XX 20-APR-2001 (first entry)

XX

DE Human heparanase, hnhp1 pn5 form, protein sequence.

XX

XX Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;

KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;

KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;

XX gene therapy; human.

OS Homo sapiens.

XX

PN WO200100643-A2.

XX

XX 04-JAN-2001.

XX

PF 19-JUN-2000; 2000WO-IL00358.

XX

XX 25-JUN-1999; 99US-0140801.

XX

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX

XX Pecker I, Michel I, Itzhaki H;

PI

DR WPI; 2001-137930/14.

DR N-PSDB; AAA91099.

XX

XX New polynucleotides and polypeptides that are distantly homologous to

PT heparanase, useful in wound healing, as well as in gene therapy

PT protocols for angiogenesis, restenosis, atherosclerosis, or

PT inflammation -

XX

XX Claim 10; Page 63; 67pp; English.

XX

CC This sequence represents a heparanase of the invention.

CC The heparanase DNA and protein sequences are useful in wound healing,

CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary

CC diseases, neurodegenerative diseases (such as Scrape, Alzheimer's

CC disease, and Creutzfeldt-Jakob disease) or viral infections. The

CC heparanase coding sequence is particularly useful in gene therapy.

XX

XX Sequence 480 AA;

Query Match 85.6%; Score 1822; DB 22; Length 480;

Best Local Similarity 86.7%; Pred. No. 5.5e-187;

Matches 351; Conservative 0; Mismatches 0; Indels 54; Gaps 1;

QY 1 NPAKSRGGGPDYLLKNYEDDIDVRSVALDKQKGCKIAQHDPVMLELQREKAAQMHVLVL 60

DB 130 npaksggpgpdyllknedy----- 149

QY 61 KEQFSNTYSNLITLTPNNYRTMHGRAVNSQLGKDYIOLKSLLOPIRIYSRASLYGPNIG 120

DB 150 -----epnnryrtmhgravnsgqlgkdyiqklsllqpiriyraslygpnig 195

QY 121 RPRKNVIALLDGFMKVAGSTDAVTWQHICYIDGRVVKVMDFLKRLDPTLSQIRKIOKV 180

DB 196 rprknvialldgfmkvagstdvatwqhcyidgrvvkvmdfklrldtldsqirkiqkv 255

QY 181 VNTYTPGKKIWLEGVVTTSAGTNNLSDSYAAGFLWNTLGLMANGQIDVIRHSHFFDHG 240

DB 256 vntytpgkkiwlegvvttsaggtnnlsdsyaagflwntlglmanqgidvirshffdhg 315

QY 241 YNHLVDQNFNPLPDYWSLLYKRLIGPKVLAVHVAGLQRPGRVIRDKLRIYAHCTNH 300

DB 316 ynhlvdqnfnpldywslllykrlgpkvlavhvaglqrpgrvirdklriyahctnh 375

QY 301 HHNNYVSGSTITFIINLHRSRKKIKIAGTLRDKLVHQYLLQPYGQGLKSKSVQLNGQPL 360

DB 376 hnnnyvrgstifilnlhrrskkiklagtlrdklvhqyllqpygqglksksvqlngqpl 435

QY 361 VMVDDGTLPDLPRPLRAGRTLVIPTVTMGFFVVKVNNALACRYR 405

DB 436 vmvddgtlpeklprlragrtlivpptvtmgffvkvnnalacryr 480

RESULT 12

AA999905

ID AM99905 standard; Protein; 214 AA.  
XX AC AM99905;  
XX DT 07-JAN-2002 (first entry)  
XX DE Human excretory related polypeptide SEQ ID NO 642.  
XX KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antischlicking; antianaemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;  
KW excretory system.  
XX OS Homo sapiens.  
XX PN WO200155313-A2.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01323.  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246509.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.

PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX  
XX WPI: 2001-465569/50.  
DR N-PSDB; AAI98878.  
XX  
XX  
XX Isolated nucleic acid molecule encoding excretory system antigen is  
PT used in preventing, treating or ameliorating a medical condition -  
PT  
XX  
XX Claim 11; SEQ ID NO 642; 574pp + Sequence Listing; English.  
XX  
XX The invention relates to novel excretory system related human  
CC polynucleotides (AAI98567-AAI99503) and the encoded proteins  
CC (AAM99594-AAM99913) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy, especially  
CC disorders related to the excretory system. The genes are isolated  
CC from a range of human tissues disclosed in the specification. The  
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 214 AA;  
SQ

Query Match 52.2%; Score 1112; DB 22; Length 214;  
Best Local Similarity 100.0%; Pred. No. 4.5e-111;  
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 VTTSAGTNNLSDSYAAGFLWNLTLGMLANQGDIVVIRHSFDDHGYNHLVDQNFNPLPDY 255  
Db 5 vttsaggtnnlsdsyaagflwntlglmianqgdvrvirhsfddhgynhlvdqnfnpdpdy 64  
QY 256 WLSLLYKRLIGPKVLAVHAGLQKRPGRVIRDKLRIYAHCTNHNHNHNVVRCSTLFI 315  
Db 65 wlslllykrligpkvlavhaglrkprgvrirdkrlriyahctnhnhnhvrgsitflfi 124  
QY 316 NLHRSRKKIKLACTLRDKLVHVLQPYGQEGUKSKSVOLNGQPLVMVDGTLPELKP 375  
Db 125 nlhrsrrkkiklactlrdklvhvlqpygqegulksksvqlngqplvmvddgtlpeikp 184  
QY 376 LRAGRTLIVPPVTMGFFVKNVNALACRYR 405  
Db 185 lragrtlivppvmgffvknvnalacryr 214

RESULT 13  
AAM43704  
ID AAM43704 standard; Protein; 214 AA.  
XX  
XX AAM43704;  
AC  
XX  
XX 24-OCT-2001 (first entry)  
XX  
XX Human bladder antigen, SEQ ID NO: 98.  
XX  
XX Human; bladder antigen; cytostatic; immunosuppressive; nootropic;  
KW neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;  
KW antiinflammatory; antitumor; antitumor; antitumor; antitumor;  
KW antifungal; antiparasitic; cardiac; gene therapy; cancer;  
KW immune disorder; cardiovascular disorder; wound healing; infection;  
KW neurological disease.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200159064-A2.  
PN  
XX  
XX 16-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US01342.  
PF  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR  
XX 04-FEB-2000; 2000US-0180628.  
PR  
XX 24-FEB-2000; 2000US-0184664.  
PR  
XX 02-MAR-2000; 2000US-0186350.  
PR  
XX 16-MAR-2000; 2000US-0189874.  
PR  
XX 17-MAR-2000; 2000US-0190076.  
PR  
XX 18-APR-2000; 2000US-0198123.  
PR  
XX 19-MAY-2000; 2000US-0205515.  
PR  
XX 07-JUN-2000; 2000US-0209467.  
PR  
XX 28-JUN-2000; 2000US-0214886.  
PR  
XX 30-JUN-2000; 2000US-0215135.  
PR  
XX 07-JUL-2000; 2000US-0216647.  
PR  
XX 11-JUL-2000; 2000US-0216880.  
PR  
XX 11-JUL-2000; 2000US-0217487.  
PR  
XX 14-JUL-2000; 2000US-0217496.  
PR  
XX 26-JUL-2000; 2000US-0218290.  
PR  
XX 26-JUL-2000; 2000US-0220963.  
PR  
XX 26-JUL-2000; 2000US-0220964.  
PR  
XX 14-AUG-2000; 2000US-0224518.  
PR  
XX 14-AUG-2000; 2000US-0224519.  
PR  
XX 14-AUG-2000; 2000US-0225213.  
PR  
XX 14-AUG-2000; 2000US-0225214.  
PR  
XX 14-AUG-2000; 2000US-0225266.  
PR  
XX 14-AUG-2000; 2000US-0225267.  
PR  
XX 14-AUG-2000; 2000US-0225268.  
PR  
XX 14-AUG-2000; 2000US-0225270.  
PR  
XX 14-AUG-2000; 2000US-0225477.  
PR  
XX 14-AUG-2000; 2000US-0225757.  
PR  
XX 14-AUG-2000; 2000US-0225758.  
PR  
XX 14-AUG-2000; 2000US-0225759.  
PR  
XX 18-AUG-2000; 2000US-0226279.  
PR  
XX 22-AUG-2000; 2000US-0226681.  
PR  
XX 22-AUG-2000; 2000US-0226688.  
PR  
XX 22-AUG-2000; 2000US-0227182.  
PR  
XX 23-AUG-2000; 2000US-0227009.  
PR  
XX 30-AUG-2000; 2000US-0228924.  
PR  
XX 01-SEP-2000; 2000US-0229287.  
PR  
XX 01-SEP-2000; 2000US-0229343.  
PR  
XX 01-SEP-2000; 2000US-0229344.  
PR  
XX 01-SEP-2000; 2000US-0229345.  
PR  
XX 05-SEP-2000; 2000US-0229509.  
PR  
XX 05-SEP-2000; 2000US-0229513.  
PR  
XX 06-SEP-2000; 2000US-0230437.  
PR  
XX 06-SEP-2000; 2000US-0230438.  
PR  
XX 08-SEP-2000; 2000US-0231242.  
PR  
XX 08-SEP-2000; 2000US-0231243.  
PR  
XX 08-SEP-2000; 2000US-0231244.  
PR  
XX 08-SEP-2000; 2000US-0231413.



Db	185	lrgrtllvippvtmgtffvkvnnalacr	214
Db	181	tsaygggapllsdtfaagfmwldklglseqlgievmrqrqffgagnyhivdenfeplpd	240
Qy	255	YWLSLYKRLIGPKVLAVHVGALQRPGRVIRDKLRIYAHCTNHHNNHNVYVGSITLFI	314
Db	241	ywlslflfkklvgpkvlmsrvkpgd-----rsklrvylhctnvhyhpryregdltlyv	291
Qy	315	INLHRSRKKIKLAGTLDKLVHQLQPYGQEGGLKSKSVQLNGOPLVMVDGTLPELKPR	374
Db	292	lnlhvntkhlkppmfsrpdvdkyllkpfsgdglksvqlngqtklmvdeqtlpaltek	351
Qy	375	PLRAGRTLVIPTVMGFFVKNVNALAC	402
Db	352	plpagsslsvpafsygffvvrnakiaac	379
RESULT	15		
AA02345			
ID	AA02345	standard; Protein; 543 AA.	
XX	XX	AA02345;	
DT	09-JUL-1999	(first entry)	
DE	XX	A human heparanase protein.	
XX	XX	Heparanase; hp; modulator; heparin-binding growth factor;	
KW	KW	cellular response; cytokine; cell interaction; plasma lipoprotein;	
KW	KW	cellular susceptibility; infection; disintegration;	
KW	KW	neurodegenerative plaque; wound healing; angiogenesis; restenosis;	
KW	KW	atherosclerosis; inflammation; neurodegenerative disease; neutralise;	
KW	KW	plasma heparin; micrometastasis; autoimmune lesion; renal failure.	
OS	OS	Homo sapiens.	
XX	XX	WO9911798-A1.	
XX	11-MAR-1999.		
XX	31-AUG-1998;	98WO-US17954.	
XX	02-JUL-1998;	98US-0109386.	
PR	02-SEP-1997;	97US-0922170.	
XX	(FRIE/) FRIEDMAN M M.		
PA	(HADA-) HADASIT MEDICAL RES SERVICES & DEV.		
PA	(INSI-) INSIGHT STRATEGY & MARKETING LTD.		
XX	Feinstein E, Pecker I, Vlodavsky I;		
PI	WPI; 1999-302255/25.		
XX	N-PSDB; AAX35648.		
XX	New human polynucleotide useful for treating angiogenesis,		
PT	restenosis, and inflammation		
XX	Claim 6; Fig 1; 63pp; English.		
PS	The specification describes a polypeptide having heparanase (hp)		
XX	activity. The recombinant protein is used as a modulator of		
CC	heparin-binding growth factors, cellular responses to heparin-binding		
CC	growth factors and cytokines, cell interaction with plasma lipoproteins,		
CC	cellular susceptibility to viral, protozoal and bacterial infections		
CC	or disintegration of neurodegenerative plaques. Heparanase may be		
CC	useful for conditions such as wound healing, angiogenesis, restenosis,		
CC	atherosclerosis, inflammation, neurodegenerative diseases, and viral		
CC	infections. Mammalian heparanase can be used to neutralize plasma		
CC	heparin, and anti-heparanase antibodies may be applied for		
CC	immunodetection and diagnosis of micrometastases, autoimmune lesions,		
CC	and renal failure in biopsy specimens, plasma samples, and body fluids.		
CC	The present sequence represents human heparanase.		
XX	Sequence	543 AA;	
SQ			

Db	185	lrgrtllvippvtmgtffvkvnnalacr	214
RESULT	14		
AA017085			
ID	AA017085	standard; Protein; 380 AA.	
XX	XX	AA017085;	
DT	21-JUL-1999	(first entry)	
DE	XX	Rat heparanase enzyme.	
XX	XX	Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;	
KW	KW	metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;	
KW	KW	arteriosclerosis; atherosclerosis; inflammation; tissue development;	
KW	KW	rat; HSPG.	
OS	OS	Rattus sp.	
XX	XX	WO9921975-A1.	
PN	06-MAY-1999.		
PD	XX		
XX	28-OCT-1998;	98WO-AU00898.	
PF	XX		
XX	09-DEC-1997;	97AU-0000812.	
PR	28-OCT-1997;	97AU-0000062.	
XX	(AUSU ) UNIV AUSTRALIAN NAT.		
PA	Freeman CG, Hamdorf BJ, Hulett MD, Parish CR;		
PI	WPI; 1999-312956/26.		
DR	N-PSDB; AAX37262.		
XX	Polynucleotides encoding mammalian endoglucuronidases, especially		
PT	heparanases, useful to promote wound healing		
XX	Claim 6; Page 87-90; 112pp; English.		
PS	The invention relates to nucleic acid sequences that encode heparanase		
CC	enzymes having endoglucuronidase activity. Recombinant heparanases are		
CC	capable of removing the HS side chain from heparan sulfate proteoglycan		
CC	(HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to		
CC	inhibit heparanase, this is useful for treatment of a physiological or		
CC	medical condition associated with elevated heparanase activity, such as		
CC	metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,		
CC	arteriosclerosis, atherosclerosis and inflammation. The human, murine and		
CC	rat heparanases can be used to enhance wound healing, especially		
CC	associated with tissue development and repair. The conditions mentioned		
CC	above can be diagnosed using specific antibodies, and also using primers		
CC	and probes specific for the heparanase polynucleotides. Other uses of the		
CC	heparanases include sequencing sulfated molecules such as HSPG. The		
CC	present sequence represents a rat heparanase.		
XX	Sequence	380 AA;	
SQ			

Query Match	37.8%;	Score	805;	DB	20;	Length	380;
Best Local Similarity	47.3%;	Pred. No.	1.4e-77;				
Matches	155;	Conservative	57;	Mismatches	106;	Indels	10;
Gaps	2;						
Qy	75	EPNNYRTMHGRVNSQLGKDYQLKSLLOPIRYSRASLYGNIGRPNKKNVIALDGFPM	134				
Db	62	epnsfwwkhisidgldgedfvelhklkq-safgnaklygpdigprgktvklrsfl	120				
Qy	135	KVAGSTVDATWQHQCIDGRVKKVMDFLKTRLDLTSLDQIRKIQKVVNTYTPGKKTWLEG	194				
Db	121	kagevidstlwhhyingrvatkedflssdvldtflsvqklkvtkmtpgkkvwlge	180				
Qy	195	VVTSAGGTNNLSDSYAAGFLWNTLGMLANQIGDVVIRHSFFDGHYNHLVDQFNPLPD	254				

Query Match	37.6%;	Score	799.5;	DB	20;	Length	543;
Best Local Similarity	38.1%;	Pred.	No. 9.6e-77;				
Matches	180;	Conservative	65;	Mismatches	139;	Indels	89;
Gaps	10;						
Qy	1	NPAKSR-CGPGPDXYL-----KNYEDDIIVRSVDALDKQGCKIAQ-HPDVMLEL	47				
Db	88	spaylrfggtktdflifdpkkestfeersyqwsqvngdi-----ckysipdpveekl	140				
Qy	48	QREKAAQHMLVLLKEQF-----SNTYS-----NULI-----	73				
Db	141	rlewpypqeql-llehyykdkfnstysrsvdvlytfancsgldlifglnallrtadlqw	199				
Qy	74	-----TEPNRYRTMHGRAVNGSQLGDYIQLKSLQLPIRIY	109				
Db	200	nssnaqlillydcsskgyniswelgnepnsofllkakdifingsqlgeqlklrk-stf	258				
Qy	110	SRASLYGPNIGRPRKNVTALLDGPWKAAGSVTDATWQHCHVIDGRVKVMDFLTRLLDT	169				
Db	259	knakiygdvqgrkrktakmlksfkagdevidsvwhhyylngtratrefrdlpnvldi	318				
Qy	170	LSDQIRIKTVKVNTVPCKKIWLGVWTTSGAGTNLSDSYAAGFLWTNLGLMANOGID	229				
Db	319	fissvqkvfvvestprgkqkwlgetsaygggappilsdtfaagfmwldkglgsarmgle	378				
Qy	230	VVIHSFFDHGYNHLDGNFNPLDPDYWLISLAYKRILGPKVLAVHAGLQRPRGVRIRD	289				
Db	379	vvmrqvfagagnyhlvdenfdpdpdywlslilffklivgtkvlmasvqgskrr-----	429				
Qy	290	KLRVYAHTCNHHNNHYVRGSTLIIFIINLHRGRKKTKLAGTRDKLVHOYLQYPXQEGL	349				
Db	430	Klrvyhctndtnprykegdltiyainlhnnvtkyrlpypfnsnqvdxyllrpgphgll	489				
Qy	350	SKSQVLNCQPLVMDDGTLPCLKPRPLRAGRTLVIPPTVMGFVFVKVNVALAC	402				
Db	490	sksqvqingitlkmwdqdtlplmekpirlprgdsqslgpatsysffvirnakvaac	542				

Search completed: July 30, 2002, 08:14:03  
Job time: 340 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:16:29 ; Search time 71.68 seconds  
(without alignments)

542.915 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_130\_534

Perfect score: 2129

Sequence: 1 NPAKSRGGPGPDYLYKNYED.....PVTMGFFVKNVNALACRYR 405

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_71.\*

1: p1r1.\*

2: p1r2.\*

3: p1r3.\*

4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1822	85.6	480	2 JC7506	heparanase protein
2	301	14.1	521	2 T45608	hypothetical prote
3	155	7.3	190	2 T01953	hypothetical prote
4	102.5	4.8	1260	2 T14022	reverse transcript
5	101	4.7	535	2 S58740	cytochrome-c oxida
6	100.5	4.7	630	2 C71374	probable glucose i
7	98.5	4.6	335	2 E84992	tryptophan--trna l
8	97	4.6	1020	2 T18342	glutamate dehydrog
9	96	4.5	898	2 B84471	cytoplasmic aconit
10	95	4.5	456	2 T24442	hypothetical prote
11	95	4.5	512	2 T11261	cytochrome-c oxida
12	95	4.5	557	2 D7210	extracellular neut
13	95	4.5	740	2 T51619	probable ethylene
14	94.5	4.4	460	2 G90554	p46-like (mycoplas
15	94	4.4	625	2 G86855	glucose inhibited
16	94	4.4	1517	2 B81393	DNA-directed RNA p
17	93	4.4	320	2 D90478	conserved hypothet
18	92.5	4.3	328	2 AC1210	drtp-D-glucose 4,6
19	92.5	4.3	876	2 G89952	DNA polymerase I
20	92.5	4.3	1616	2 T16600	vitellogenin vit-1
21	92.5	4.3	2241	2 S09811	hypothetical prote
22	92	4.3	455	2 S67627	probable membrane
23	92	4.3	461	1 S65187	GPI-anchor biosynt
24	92	4.3	534	2 S17993	cytochrome-c oxida
25	92	4.3	898	2 T10101	aconitate hydratase
26	91.5	4.3	986	2 F98229	sarcosine oxidase
27	91.5	4.3	986	2 A83056	hypothetical prote
28	91	4.3	1272	2 C90593	probable lipoprote
29	90.5	4.3	283	2 H81435	

30	90.5	4.3	682	2 A44493	serum-inducible ki
31	90	4.2	650	2 F81654	regulatory protein
32	90	4.2	709	2 T16584	hypothetical prote
33	89	4.2	687	2 A50345	polyphosphate kina
34	89	4.2	770	2 S77523	hypothetical prote
35	89	4.2	2380	2 E71604	hypothetical prote
36	88.5	4.2	286	2 A86576	chromosome partiti
37	88.5	4.2	286	2 B72047	probable fatty-acy
38	88.5	4.2	341	2 D70761	ATP-dependent heli
39	88.5	4.2	933	2 H90247	PRP8 protein homol
40	88.5	4.2	2403	2 T30875	cytochrome c oxida
41	88	4.1	555	2 A52147	hypothetical prote
42	88	4.1	1032	2 S53571	calcium channel pr
43	88	4.1	1106	1 CHRBA2	collagen alpha 3(V
44	88	4.1	3137	1 A37797	hypothetical prote
45	87.5	4.1	411	2 D87377	

ALIGNMENTS

RESULT 1  
JC7506  
heparanase protein 2a - human  
C:Species: Homo sapiens (man)  
C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 01-Dec-2000  
C:Accession: JC7506  
R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hirccock, M.;  
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000  
A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase famli  
A:Reference number: JC7506  
A:Accession: JC7506  
A:Molecule type: mRNA  
A:Residues: 1-480 <MCK>  
A:Cross-references: GB:AF282885  
C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and th  
therapies.  
C:Genetics:  
A:Gene: hpa2a  
A:Map position: 10q23-10q24  
C:KeyWords: heparin binding; membrane bound





C:Superfamily: tryptophan--trna ligase  
C:Keywords: ligase

Query Match 4.6%; Score 98.5; DB 2; Length 335;  
Best Local Similarity 21.1%; Pred. No. 2.5;  
Matches 50; Conservative 38; Mismatches 78; Indels 71; Gaps 10;

Qy 22 IVRSDVALDKQKCKIAQHPDVMLELOREKAAQMHVLVLLKEQFSNTYSNLILTEPNNVRT 81  
Db 132 LMAADILLYQTNPVFPQDQKHVELTRN-----IAHRFNSLYGH-VFTLPKPLIT 181

Qy 82 MGRVNSQLGKDKYIQLKSLLOPIRIYSRASLYCPNTPGRPKNVIALLDG-----FMKVA 137  
Db 182 OHGS-----KIMSLEPSKMSKSDI-----NKNNVIFLLDDIKTVISKIQ 222

Qy 138 GSTVDVAVTWHQCYID-----GRVVKVM-----DFL-----KTRLLDTL 170  
Db 223 NAYTSETPSKIYYDIEKKPGISNLLEILSAITNKDIDILLKEGLMYSEFKNIVADHL 282

Qy 171 SQQIRKIQRVNTYTPG-----KKIWLEGVVTTSAGTNNLSDSYAAGFLWLTGLML 223  
Db 283 SKFLYKLOKSYNDRNDEVYLLKIAIEGAMKSLKSNKTLTKVY-----DKLGLI 332

RESULT 8  
T18342  
glutamate dehydrogenase (EC 1.4.1.2) precursor - Sauroleishmania tarentolae  
C:Species: Sauroleishmania tarentolae  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-May-2000  
C:Accession: T18342  
R:Brigaund, F.; Stripecke, R.; Frech, G.C.; Freedland, S.; Turck, C.; Byrne, E.M.; Simps  
Mol. Cell. Biol. 17, 3915-3923, 1997  
A:Title: Mitochondrial glutamate dehydrogenase from Leishmania tarentolae is a guide RNA  
A:Reference number: Z18878; MUID:97342629  
A:Accession: T18342  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-1020 <BR>  
A:Cross-references: EMBL:U31177; NID:945092; PID:945093; PIDN:AAB62735.1  
C:Gene: GDH  
C:Superfamily: glutamate dehydrogenase  
C:Keywords: NAD; oxidoreductase

Query Match 4.6%; Score 97; DB 2; Length 1020;  
Best Local Similarity 26.6%; Pred. No. 16;  
Matches 38; Conservative 26; Mismatches 45; Indels 34; Gaps 6;

Qy 7 GGP-----GPDYILKNYEDDITVRSDVA-----LDKQKCKIAQHPDVMLELOREKAAQ 54  
Db 670 GGPDDGLGSNEVLSRSKKVMQVMDISASLHDPNGIDREELARLAHRLPLREFSRK--- 726

Qy 55 MHLVLLKEQFSNTYSNLILTEPNNYRTMHGRAV-NSQLGKDYIQLKSLLOPIRIYSRAS 113  
Db 727 ----LSPEGE-----LVLTEDHNKVLPGDTLVEDGSRLNREFHLK-----YSAD 768

Qy 114 LYGNIGRPRKNVIALLDGFMKV 136  
Db 769 VFVPCGGRPRSVLTENVGRELKV 791

RESULT 9  
B84471  
cytoplasmic aconitate hydratase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: B84471  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: B84471  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-898 <STO>  
A:Cross-references: GB:AE002093; NID:g4586021; PIDN:AAD25640.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: AL2g05710  
A:Map position: 2  
C:Superfamily: iron-responsive element-binding protein

Query Match 4.5%; Score 96; DB 2; Length 898;  
Best Local Similarity 20.3%; Pred. No. 16;  
Matches 98; Conservative 61; Mismatches 180; Indels 144; Gaps 22;

Qy 20 DDIVRSDVALDKQKCKIAQHPDVMLELOREKAAQMHVLVLLKEQF-----SNTYSNLIL 73  
Db 129 DHSQVDVARSN-----AVQANMELEFQRN-----KERFAFLKMGSTAFQNNLV 173

Qy 74 TEPNRYRTMHGRAVNSQLGKDYIQLKSLLOPIRIYSRAS-----L 114  
Db 174 VPPGS-GIVH--QVNLEYLGRVVENTKGLLYPDSVGTDSHTTMDGLGVAGVGGIEA 230

Qy 115 YGPNIGRPRKNVIALLDGFMKVAGSTVDVAVTWHQCYIDGRVVKVMDFLKTLLDLSQDI 174  
Db 231 EATMLGQPMNSVLPGVVG-F-KLAGKMRNGVT-----ATDLVLTQML 272

Qy 175 RK--IQKVNTYTPGKKIWLLEGVVTTSAGTNNLSDSYA--GFLWLN--TLGMLANQG 227  
Db 273 RKHGVGVGEVFEYFNG-----MSGLSLADRYATNNSPEYATMGFFPVDHVTLYLKLGTG 328

Qy 228 -----IDVVIR--HSFEHDG-----YNHLVDQNFNPLDY----- 255  
Db 329 RSETVAMIEAYLRANMNFVDYNEPQQDRVYSSYLELDDVEPCISGPKRPHDRVTLKE 388

Qy 256 ----WLSLL-----YKRLIGPKVLAVHVAGLQRPGRVIRDKLRIYA--HCTNHHNHN 304  
Db 389 KMAWHSCLDKSKVGKFAIPKEAQEVVNFSDGQPAELKHGSVVIAAITSTNTSNPS 448

Qy 305 YVRS--ITLFIINLHRSRK--IKLAGTLRDKLVHQLQPYQVQGLKSKSVQLNGQ--P 359  
Db 449 VMLGAGLVAKKACDGLQVQKPIKTSAPGSGVVTYLLKSLQLEYLNEQGENIVGYGCT 508

Qy 360 LVMVDDGTLPDL-----KPRPLRAGRITLVIPTVTMGFFVVK 395  
Db 509 TCIGNSGEINESVGAAITENDIVAAVLSGNRNFEGRVHPLTRANYLASPLVAVYALAG 568

Qy 396 NVN 398  
Db 569 TVN 571

RESULT 10  
T2442  
hypothetical protein T04B2.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T2442; T26042  
R:Coles, L.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: Z19890  
A:Accession: T2442  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-456 <WIL>  
A:Cross-references: EMBL:Z68299; PIDN:CAA92615.1; GSPDB:GN00022; CESP:T04B2.5  
R:Coles, L.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: Z20141  
A:Accession: T26042



QY 341 QPYGEGELKSKSVOLNGQPLVMVDGT-LPELKPRPLR--AGRTIVIPVTMGFFV 393  
Db 413 QTYDYFNKNGNWAFSASDWVVGDLFLNNTTHRALSLANPTLYDQPDNNNTV 468

RESULT 13  
T51619  
probable ethylene receptor [imported] - netted muskmelon  
C:Species: Cucumis melo var. reticulatus (netted muskmelon)  
C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 08-Dec-2000  
C:Accession: T51619  
R:Sato-Nara, K.; Yuhashi, K.I.; Higashi, K.; Hosoya, K.; Kubota, M.; Ezura, H.  
Plant Physiol. 120, 321-330, 1999  
A:Title: Stage- and tissue-specific expression of ethylene receptor homolog genes during  
A:Reference number: 225417; MUID: 99252277  
A:Accession: T51619  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-740 <SAT>  
A:Cross-references: EMBL:AF054806; PIDN:AAC99645.1  
C:Genetics:  
A:Gene: ETR1  
C:Superfamily: response regulator homology

Query Match 4.5%; Score 95; DB 2; Length 740;  
Best Local Similarity 18.9%; Pred. No. 15;  
Matches 87; Conservative 79; Mismatches 143; Indels 152; Gaps 23;

QY 20 DQIVRSVALD---KQKCKIAQHPVMELEQREKAAQMHLY-----LKQFSNTYSNL 71  
Db 321 DLLMEQNALDLARAEATAIRANDFLAVMNHMRTPMHAIALLSLLOQTELTPEQRL 380

QY 72 ----ILTEPNNTYTHMGRVAVSGQKDYIQLK-----SLQPIRYSRAS 113  
Db 381 MVETILKSSNLLATLNDVLDLSRLDEGSLQLDIGTFNLHAFVKEVLNLIKPVTLVKKLS 440

QY 114 L---YGNP-----IGRPKNVIALI-----DGFMKVAGSTVDVAVTQHCVIDGR 154  
Db 441 LTLHGPDLVPFVAGDEKRLMAQLNVGNVAVKFSKESISAIYAKSETF-----R 493

QY 155 VYKVMDF-----LKRLLDRLS-----DQIRKIQVNVNTYTPGKKIWLEGVVTTSA 200  
Db 494 EIRVPDFHPVPSDSHYLRVQVKDTGSGISPDIPKLTFFKAQTVGPR-----NS 544

QY 201 GGTN-----NLSDSYAGFLWNTLGLMANGIDVIRHSFFDHGYNHLVDQNFN 250  
Db 545 GSGGLGLATCKRFVNLME----GHWLESEGLGKGTATFIVKLGTD-----QSNES 593

QY 251 PLPDVWLSLLYKRLI-----GPKVLAVHVAGLQRPGRVIRDKLRIYAHCTNHHNNY 305  
Db 594 KLP--YTSKIHEHSHTSPFGKLVLMVDNGVRSRVTKG-----LLVHLGC----- 637

QY 306 VRGSITFLTINLHRSRKKIKLAGTLRD-----KLHVQYLLQPYQGEQ-----LKS 350  
Db 638 -----EVTAGSIEEFLRVVVSQEHKVVFMFDICTP-GVDGYELAIRIRE 679

QY 351 KSVQLNGQPLVMVDGTLPEL-KPRPLRAGR-LVIPPVTM 389  
Db 680 KFAKCHERPFMVYLTGNSDKVTKESCLRAGMDGLILKPVSI 720

RESULT 14  
G90554  
p46-like (mycoplasma hyorhinis) lipoprotein [imported] - Mycoplasma pulmonis (strain UAE  
C:Species: Mycoplasma pulmonis  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001  
C:Accession: G90554  
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;  
Nucleic Acids Res. 29, 2145-2153, 2001  
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm  
A:Reference number: A99512; MUID: 21267165; PMID:11353084

A:Accession: G90554  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-460 <KUR>  
A:Cross-references: GB:AL445566; PID:gl14089757; PIDN:CAC13516.1; GSPDB:GN00153  
A:Experimental source: strain UAB CTIP  
C:Genetics:  
A:Gene: MYPUL\_3430  
A:Genetic code: SGC3

Query Match 4.4%; Score 94.5; DB 2; Length 460;  
Best Local Similarity 19.6%; Pred. No. 8.3;  
Matches 78; Conservative 53; Mismatches 145; Indels 121; Gaps 18;

QY 60 LKEQFSNTYSNLLTEPNNTYTHMGRVAVSGQKDYIQLKSLLOPIRYSRASLYGPN 119  
Db 55 VSEKISDGSIVITVDPENQR-----WVETKKQLD-----AYSKT----- 89

QY 120 GRPRKNVIALLDGFMKVAGSTVDVAVTQHCYIDGRVVKVMDFLKTRLLDLSQ----- 173  
Db 90 -----NGFEHSSNHVKAQEQNSFVDAELAKTGNNKPKVVLGMAADSGNATQA 138

QY 174 -----IRKIQK-----VVNTYTPGKKIWLEGVVTTSA-----GQTNNL 206  
Db 139 IBSTTNAQAQFTAVDRFTHKISSNNYVYVAPNNYTVGQ---LQGLALISGIYKGQGEPEK 195

QY 207 SDSYAAGFLWNTL-----GMLANQGI-DVVIHRSFEDHGYNHL-----VDQNFNPLPDY 256  
Db 196 TLEEAKTYVMANKLASEKGFVALGAPEDNNSHLFEKGMADVITAIMKIDSNLKYFGE 255

QY 257 LSLLYKRLIGPKVLAVHVAGLQRPGRV---IRDKLRIYAHCTNHHNNVVRGS---I 310  
Db 256 GADVNRLSG-----DSKPTDQELTSAITSHFVVAQA---NWNYSQKGQFLD 300

QY 311 TLFITLHRSRKKIKLAGTL--RDKLHVQYLLQPYQGEGLKSKSVOLNGQPLVMVDGTL 368  
Db 301 TLFKASTFDSRKN-NLAAVLAPNDEMAQAAITSEIQKGLDPKKIYITGQ-----DSNQ 353

QY 369 PELKPRPLRAGRTLVP-----PVTMGFFVVK 396  
Db 354 PSLSRIDNETGQNMITSKEDWKIAIAISTALAYIVKN 390

RESULT 15  
G86855  
glucose inhibited division protein Gida [imported] - Lactococcus lactis subsp. lactis  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C:Accession: G86855  
R:Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Eh  
Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis  
A:Reference number: A86625; MUID: 21235186; PMID:11337471  
A:Accession: G86855  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-625 <STO>  
A:Cross-references: GB:AE005176; PID:gl12724877; PIDN:AAK05945.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: gida  
C:Superfamily: gida protein

Query Match 4.4%; Score 94; DB 2; Length 625;  
Best Local Similarity 19.6%; Pred. No. 14;  
Matches 72; Conservative 58; Mismatches 119; Indels 118; Gaps 17;

QY 100 KSLLOPIRYSRASLYGPNIGRPRKNVIAL-LDGFMKVAGSTVDVAVTQHCYID---GR 154  
Db 31 KYLLMTINLNVAVFMPCNPSIGSAGIYVREIDALGGEMGRNIDKTYIQMKMLNTGKGP 90

```
QY 155 VVKVM-----DFLKTRLLDFTLSOQ-----IRKI-----QKVNTYT-PGKKIWL 192
Db 91 AVRALRAQADKDEYADSMKNTVSDQENLTLRQGMVEELILDEEKKKVIGIKTSTGKYGA 150
QY 193 EGVVTT-----SAGTNNLSDSYAAGFLWNLTLGMLANOGIDVVIRHSF 236
Db 151 KAVIITGTALRGEIIIGELKYSSGPNNLS-----SIGLADN-----LREIG 193
QY 237 FDHGYNHLVDQNFNPLPDYWLSLYKRLIGPKVLAVHVAGLQKPRPG----- 284
Db 194 FEIG-----RFGTGTPPRVLASSIDYDKTEIQPGDEAPNHFSEMS 234
QY 285 --RVIRDKLRIYACHTNHHNNHYVSGTITFIINLHRS---RKKIKLAG-----TLRDKL 334
Db 235 DENYKDOIPCWLTYYTTTENSHTILRG-----NLHRAPLFSGIVKGYPYCPYCPYCPY 287
QY 335 V-----HOYLLQPYGOEGLKSKSVOLNGOPLVWVDGTLPELKPRLRAGRTLIVIPPV 387
Db 288 TRFADKPRHQLFLEP---EGRNTEEVYIGGLSTSWPEDVQFDLVKSIPLGLENAQMMRPGY 344
QY 388 TMGFFVY 394
Db 345 AIEYDVV 351
```

Search completed: July 30, 2002, 08:16:32  
Job time: 349 sec

**THIS PAGE BLANK (USPTO)**





QY	147	QHCYIDGRVVKVMDFLKTRLLDTLSQIRKIQKVVNTYTPGKIWLGVVTTSGAGTNNL	206
DB	119	LHLQYD-----FVVDVCSNTTDAAGVYGAHAHVVT--ARGRRISARAVLTTG-----	166
QY	207	SDSYAAGFLWL-----NTGLMANQGDVIVRHSPFDHG-----YNHLVDQNF	249
DB	167	--TFMEGRVYIGEXEBAPEGRLGEHAAEGALRRKKGFQMLTKGTTPARVLRKSV-----	221
QY	250	NLPDYWLSLLYKR-----LIGPKVLAVHAGLQRPGRVIRDKLRIYAHCTNNHHNY	305
DB	222	-----LSVMEKQADAIIMRPFSA-HV-----EINRHADCYINVTNERTHQL	263
QY	306	VRGSITLFTINIHR-----RKKTKLAGT-----LRDLVHQYLLQPYQOEGLS	350
DB	264	IRE-----NHRSPFSGRIKAVGTGYPCPSIEDKVRKFPDRIRHQYIEP-----	313
QY	351	KSVALNQPLVMVDD-----GTLPELK-----PRPLRAGRLVLPVPTMGF-FVVKVNA	399
DB	314	EELYINGLSSCLPEDIODEMIRTIQMERAVITRPAYADVAVLPVQLGIDLQTKRVSG	373
QY	400	L 400	
DB	374	L 374	
RESULT	2		
SYW_BUCAI		STANDARD;	PRT; 335 AA.
ID	SYW_BUCAI		
AC	P57602;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DE	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	tryptophanyl-tRNA synthetase (EC 6.1.1.2) (tryptophan--tRNA ligase)		
DE	(TrpRS).		
GN	TRPS OR BU536.		
OS	Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum		
OS	symbolic bacterium)		
OC	Bacteria; Proteobacteria; gamma subdivision; Buchnera.		
OX	NCBI_TaxID=118099;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RP	STRAIN=TOKYO 1998;		
RX	MEDLINE=20445173; PubMed=10993077;		
RA	Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;		
RT	genome sequence of the endocellular bacterial symbiont of aphids		
RT	Buchnera sp. AFS.;		
RL	Nature 407:81-86(2000).		
CC	- - CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) -> AMP +		
CC	diphosphate + L-tryptophanyl-tRNA(Trp).		
CC	- - SUBUNIT: HOMODIMER (BY SIMILARITY).		
CC	- - SUBCELLULAR LOCATION: Cytoplasmic.		
CC	- - SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; AP001119; BAB13229.1; -		
DR	InterPro; IPR001412; trna-synt_1.		
DR	InterPro; IPR002306; trna-synt_trp.		
DR	Pfam; PF00579; trna-synt_1b; 1.		
DR	PRINTS; PR01039; TRNASYNTHTRP.		
DR	PROSITE; PS00178; AA-trna_LIGASE_I; 1.		
KW	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;		
KW	Complete proteome.		
FT	SITE	14 22	"HIGH" REGION.
FT	SITE	196 200	"KMSKS" REGION.
FT	BINDING	199 199	ATP (BY SIMILARITY).
FT	SEQUENCE	335 AA; 38549 MW; 871562D5A6734E3F	CRC64;

QY	22	IVRSVDALDKGCKGAQHPDVMLELQREKAAQMHVLVLLKEQFSNTYSNLIITPEPNYRT	81
DB	132	LMAADILLYTNTVPVQDQKQHVLTNR-----IAHRENSLYGH-VFTLPKPLIT	181
QY	82	MHGRAVNSQLGKDYTLQKSLLOPIRIYSRASYGNIGRPNKVNIALLDG-----FMKVA	137
DB	182	QHGS-----KIMSLLEPSKMKMSKSDI-----NKKNVIFLLDDIKTVISKIQ	222
QY	138	GSTVDATWQHCYID-----GRVVKVM-----DEL-----KTRLIDTL	170
DB	223	NAYTDSPTSKYIYDIEKKPGISNLNLEILSAITNKDIDILLKEGLMYSEFNIVADHL	282
QY	171	SDQIRKIQKVVNTYTPG-----KKIWLEGVVTTSAGTSNNISDSYAAGFLWLNTLGLM	223
DB	283	SKFLYKLQKSYNDYRNDEVVLKTIAYEGAMKSQLSKNKTLTKVY-----DKLGLI	332
RESULT	3		
GIDA_LACLA		STANDARD;	PRT; 625 AA.
ID	GIDA_LACLA		
AC	OQCEJ4;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Glucose inhibited division protein A.		
GN	GIDA OR LL1847.		
OS	Lactococcus lactis (subsp. lactis) (Streptococcus lactis).		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC	Lactococcus		
OX	NCBI_TaxID=1360;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RP	STRAIN=IL1403;		
RX	MEDLINE=21235186; PubMed=11337471;		
RA	Bolotin A., Wincker P., Manger S., Jaillon O., Malarme K.,		
RA	Weissenbach J., Ehrlich S.D., Sorokin A.;		
RT	"The complete genome sequence of the lactic acid bacterium Lactococcus		
RT	lactis ssp. lactis IL1403.;"		
RL	Genome Res. 11:731-753(2001).		
CC	- - FUNCTION: NOT KNOWN.		
CC	- - SIMILARITY: BELONGS TO THE GIDA FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; AE006414; AAK05945.1; -		
DR	InterPro; IPR001327; FAD_pyr_redox.		
DR	InterPro; IPR002218; GIDA.		
DR	InterPro; IPR001100; pyr_redox.		
DR	Pfam; PF01134; GIDA; 1.		
DR	PRINTS; PR00368; FADPNR.		
DR	PRINTS; PR00411; PNDRDTASEI.		
DR	ProDom; PD003738; GIDA; 1.		
DR	PROSITE; PS01280; GIDA_1; 1.		
DR	PROSITE; PS01281; GIDA_2; 1.		
KW	Complete proteome.		
QY	SEQUENCE	625 AA; 69377 MW; D79C35B395C61147	CRC64;

Query Match 4.4%; Score 94; DB 1; Length 625;  
Best Local Similarity 19.6%; Pred. No. 4.8; 119; Indels 118; Gaps 17;  
Matches 72; Conservative 58; Mismatches

```

Qy 123 RKNVIALD-----GFMKVGASTVDAVT---WQHCYIDGRVVVKWDFL-----KTRLLDTL 170
Db 607 KKDIAALVDYVYKOGGUEITASFDRUKNLGFEYATKAGISIASIADIIVPNDKOKAIDEA 666
Qy 171 SDQIRKIOKVNTYTPGKKIWLEGVVTTSAGCTNNLSDSYAAGFLWLNTLGLMLANQIDV 230
Db 667 KKQVREIQ---NSYNLG-----LITSGRYNKIID-----IWKSTNNVLSEMMKL 709
Qy 231 VIRHSFFDHGYNHLVDQNFNLPDYWLSLYKRLIGPKVLAVHVA-----GLQKRPKP 284
Db 710 VEKOK---EGEN-----SIYMWADSGARSAQISQLAAMRGLMTPK-DG 750
Qy 285 RVIRDKLRIYAHCTNNHHNNHYVRG-SITFLTIINLHRSRK-----KTKLAGTIRLDKLIVH 336
Db 751 STIEPTI-----TSNREGINVLVEYFTSITHGARKGLADYALKTANAGYVYTRKILID 800

```

```

01 QY 337 -----QYLQPTGEGLEKSKSVQLNGQQVAVDGTLTLPRLPRLRAGRTIL---VIPPVV 388
      :   :       :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db    801 VAQNKKITIEDCGTH-----EGVEINE-----ITADSSIIETLEERIL--GRVLAEVDIDPIT 851
      :   :       :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY    389 -----MGFFVVKNVN---ALACR 403
      :   :       :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db    852 NSVFAEGLTMDERKAKILGESGIKSVMIRPTICK 887
      :   :       :   :   :   :   :   :   :   :   :   :   :   :   :   :

RESULT          5
LKTA_PASSP
ID LKTA_PASSP STANDARD; PRT; 947 AA.
AC P55123;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Leukotoxin.
GN LKTA.
OS Pasteurella haemolytica-like sp. (strain 5943B).
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
CC Pasteurella.
OX NCBI_TaxID=28165;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93239320; PubMed=8478098;
RA Chang Y.-F., Ma D.-P., Shi J.J., Chengappa M.M.;
RT "Molecular characterization of a leukotoxin gene from a Pasteurella
RF haemolytica-like organism, encoding a new member of the RTX toxin
RL family."
RL Infect. Immun. 61:2089-2095(1993).
CC -!- FUNCTION: VIRULENCE FACTOR WHICH IS CYTOTOXIC FOR LEUKOCYTES BUT
CC IS NOT HEMOLYTIC.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
CC CALCIUM; THE GLY-RICH REGION IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC ACTIVITY.
CC -!- DOMAIN: THE THREE TRANSMEMBRANE DOMAINS ARE BELIEVED TO BE
CC INVOLVED IN PORE FORMATION BY THE CYTOTOXIN (BY SIMILARITY).
CC -!- PTM: PALMITOYLATED BY LKTC. THE TOXIN ONLY BECOMES ACTIVE WHEN MODIFIED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at

```

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: L12148; AAL16444.1; -;  
DR InterPro: IPR001343; HemLysn\_Ca\_bind.  
DR InterPro: IPR003355; RTX\_N.  
DR InterPro: IPR003995; RTX\_A.  
DR Pfam: PF00353; hemolysincabind; 1.  
DR Pfam: PF02382; RTX; 1.  
DR PRINTS: PR00313; CABNDNGRPT.  
DR PRINTS: PR01488; RTXTOXINA.  
DR PROSITE: PS00330; HEMOLYSIN\_CALCIIUM; 4.  
KW Hemolysis; Toxin; Cytolysis; Cytotoxin; Repeat; Calcium;  
KW Transmembrane; Lipoprotein; Palmitate.  
FT TRANSMEM 154 170 POTENTIAL.  
FT TRANSMEM 312 333 POTENTIAL.  
FT TRANSMEM 393 414 POTENTIAL.  
FT DOMAIN 625 780 7 X REPEATS, GLY-RICH.  
FT REPEAT 625 630 1.  
FT REPEAT 625 630 2.  
FT REPEAT 730 735 1.  
FT REPEAT 739 744 3.  
FT REPEAT 748 753 4.  
FT REPEAT 757 762 5.  
FT REPEAT 766 771 6.  
FT REPEAT 775 780 7.  
SQ SEQUENCE 947 AA; 101559 MW; 9744F06395EF5BED CRC64;

Query Match 4.4%; Score 93.5; DB 1; Length 947;  
Best Local Similarity 24.3%; Pred. No. 9.3;  
Matches 64; Conservative 30; Mismatches 80; Indels 89; Gaps 13;  
QY 13 YLKNYEDDIVRSVALDKQCKTAQHPDVMLELQREK-----AAQHVLVLLKEQFSNT 67  
DB 43 YIPKDYEDSGRGQLDLVRAE-----DLGIEVOREERNGIATAQNSLSTIQNILGFS 97  
QY 68 YSNLILTEPNRYTHMGRVNGSQLGKDYIOLKSLLOPIRIYVSRLSYGPNIGRPKNVI 127  
DB 98 ERGVVLSAP-----QDLKLLQYKI-SKAPGSSENVAKNLGNAQ 135  
QY 128 ALLDGFMKVAGSTVDVATWQHGYIDGRVVK-----VMDFLKTRL-----L 167  
DB 136 TLLSGIQSLGSMAGMD-----LD-EILKNKGSELDLAKAGLELTNSLIENIANSVQTL 189  
QY 168 DTLSDOI-----RKIQVNTYTPGKK-----TWLEGVVTTSAGTNN--LSDSY 210  
DB 190 DTFSEQISQLGTLQNVKGLGLTGDKNFSGFSKAGLGLEIVISGLLSGATAALVLADKN 249  
QY 211 A-----AGELWLNTLGLMLAQ 226  
DB 250 ASTDRKVGAGF-----ELANQ 265  
RESULT 6  
SNK\_RAT  
ID SNK\_RAT STANDARD; PRT; 682 AA.  
AC Q9R012;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE serine/threonine-protein kinase SNK (EC 2.7.1.-) (Serum inducible  
DE kinase).  
GN SNK.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99452760; PubMed=10523297;

RA Kauselmann G., Weller M., Wulff P., Jessberger S., Konietzko U.,  
RA Scafidi J., StaUBL U., Bereiter-Hahn J., Streibhardt K., Kuhl D.;  
RT "The polo-like protein kinases Fnk and Snk associate with a Ca(2+)-and  
RT interin-binding protein and are regulated dynamically with synaptic  
RT plasticity";  
RT EMBO J. 18:5528-5539(1999).  
CC CC FUNCTION: MAY PLAY A ROLE IN THE DIVISION OF AT LEAST SOME CELL  
CC TYPES, SUCH AS FIBROBLASTS, AND COULD FUNCTION IN EMBRYOGENESIS,  
CC WOUND HEALING OR NEOPLASIA (BY SIMILARITY).  
CC -I- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC CDC5/POLO SUBFAMILY.  
CC -I- SIMILARITY: CONTAINS 2 POLO BOX DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AFI36583; AAF08366.1; -;  
DR HSP: P00518; IphK  
DR InterPro: IPR00719; Euk\_pkinase.  
DR InterPro: IPR000959; POLO\_box.  
DR InterPro: IPR002290; Ser\_thr\_pkinase.  
DR Pfam: PF00069; pkinase; 1.  
DR Pfam: PF00659; POLO\_box; 2.  
DR SMART: SM00220; S\_TKC; 1.  
DR PROSITE: PS00078; POLO\_BOX; 2.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS00101; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding; Repeat.  
FT DOMAIN 54 59 POLY-HIS.  
FT DOMAIN 79 331 PROTEIN KINASE.  
FT NP\_BIND 85 93 ATP (BY SIMILARITY).  
FT BINDING 108 108 ATP (BY SIMILARITY).  
FT ACT\_SITE 202 202 BY SIMILARITY.  
FT DOMAIN 507 570 POLO BOX 1.  
FT DOMAIN 603 674 POLO BOX 2.  
SQ SEQUENCE 682 AA; 77919 MW; 58C50DEBDE83D5F3 CRC64;

Query Match 4.3%; Score 92.5; DB 1; Length 682;  
Best Local Similarity 22.7%; Pred. No. 7.2;  
Matches 94; Conservative 58; Mismatches 161; Indels 101; Gaps 21;  
QY 6 RGGPGDYLYKNVEDDIVRSVALDKQCKTAQHPDVMLELQREKAAQ-----M 55  
DB 87 KGG-----FAKCYE-----MTDLTNKVVAAKLIHRSVAKPHQREKIDKEIHLHLHK 137  
QY 56 HLVLKKEQFSNTYSNLILTEPNRYTHMGRVNGSQLGKDYIOLKSLLOP-IRIYSRASL 114  
DB 138 HVQFYHYFEDKENIYILLEYCSRSM-----AHILK-----ARKVLTEPEVRYLRQIV 187  
QY 115 YGPNTGRPRKNVIALDGFMKVAGSTVDVATWQHGYIDGRVVKVWDFLKLRLDLSQI 174  
DB 188 SGLKYLHEQE-----ILHRDLKLGNNFIN-----EAMELKVGDFGLAARLEPLEHRR 234  
QY 175 RKIQKVNTYTP-----GKKIWLEGVV--TTSAG-----GTNLLSDSY-----AAGF 214  
DB 235 RTICGTPTNLSPEVLKNGHGCESDIWALGCVMYTLLGRPPFTTNLKYRCIARY 294  
QY 215 LWLNTL-----GMLAQ-----GIDVIRHRSFFDHGY--NHLVDONFNPDPYWLIS 258  
DB 295 TWPSSLLAPAKHLIASMLSKNPEDRPSLDDIIRHDFLQGGFTPDRLSSSCCHTVPDFHLS 354  
QY 259 LLYKLLGKPKVLAVHVGQRKPR-----PGRVIRDKLRIYACTNHHNHHNHYRGSTILFI 314  
DB 355 SPANKFF-KAAAAALFGGKKDKARYNDTHNKVSKDEDIY-----KLRLDKAKTSITQ-Q 407  
QY 315 INLHRSRKKIKLAGTLRDLKLVHQLLYLPQYQSGELKSKSVQLNGQPLVMVDDGTL 368



SQ SEQUENCE 1616 AA; 188063 MW; F173D7C452D123F5 CRC64;  
 Query Match 4.3%; Score 92.5; DB 1; Length 1616;  
 Best Local Similarity 19.7%; Pred: No. 23;  
 Matches 83; Conservative 69; Mismatches 141; Indels 129; Gaps 19;

QY	17	NYEDDIVRSDVA--LDKQKCKIAQHPDVML-----EIQREKAQAQHLVLK 61
DB	205	NFDKCIERSEIAVGLRYSSECEPECKDVLIRPQTWYVILENEELKESEVRSLYTNVN 264
QY	62	EQ---PSNTYSNLIITPEPNYRTMHGRAVNSQLGKDYIQLKSLLOPIRIYSRASLYGN 118
DB	265	GQVYMKTETRSKVLLENHSIKS-HIEKYNGE---KESI-----IYS----- 302
QY	119	IGRPKNVIALLDGFMKVAAGTVDVATWTOHCYIDGRVVKVMDFLTRLDTLSDQIRKI 178
DB	303	-----SRWEQLVEDFEK-NGDKAEAFPEKFFLD-----KKMLIKTITEIQEVE 347
QY	179	KVNVTYPGKKIIEGVV-----TTSAGTNNISDSYAAGFLMNTLGLMANOGIDVIRH 234
DB	348	NNI-----PETSHEFLARIVRIFTTS---TSQKEIH-----ETLYVKADKKIOSLMEH 393
QY	235	SF-----FDHCYNHLVDQENPLPDYWLSSLYKRLIGPKVLAVHVAGLQKRPGRV 286
DB	394	ALATAGTKNTYIQTILVHIENEDIVPLE-----AAQLLSKIQTETPPSQT 437
QY	287	IRDKLRIYAHCTNNHNNHYR-----GSITFIINLHRSR----- 321
DB	438	IAEALIAFAESRVSKNQVYRQSAWLAAGSVVRGIVDYKNIRPLVREDKRELKEPLRVF 497
QY	322	-KKTKLAGTRDKLVHQYLLQPYGQEGKSKSVQGLVWVDDGTLPKPLRPLRAGR 380
DB	498	MQQYKDAETTYEKL---ALKSIGNAGLIDISVQNLNE---IIVDKQLLPVRKEAIDALR 551
QY	381	TL 382
DB	552	LL 553

RESULT 9  
 TEGU\_HCMVA STANDARD; PRT; 2241 AA.  
 AC P16785;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-APR-1990 (Rel. 15, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE Probable large tegument protein.  
 GN UL48.  
 OS Human cytomegalovirus (strain AD169).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Cytomegalovirus.  
 OX NCBI\_TaxID=10360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90269039; PubMed=2161319;  
 RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,  
 RA Horanell T., Hutchison C.A. III, Kouzarides T., Martignetti J.A.,  
 RA Readle E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;  
 RA "Analysis of the protein-coding content of the sequence of human  
 RT cytomegalovirus strain AD169";  
 RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).  
 CC -!- FUNCTION: TEGUMENT PROTEIN.  
 CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,  
 CC EHV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).









Search completed: July 30, 2002, 08:31:33  
Job time: 980 sec

```
RESULT 15
UVRE_PASMU
ID UVRE_PASMU STANDARD; PRT; 678 AA.
AC P57844;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Excinuclease ABC subunit B.
GN UVRB OR PM0429.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRB STIMULATES
CC THE ATPASE ACTIVITY OF UVRA IN THE PRESENCE OF UV-IRRADIATED
CC DOUBLE-STRANDED DNA. IT ALSO ENHANCES THE ABILITY OF UVRA TO BIND
CC TO UV-IRRADIATED DUPLEX DNA (BY SIMILARITY).
CC -!- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE UVRB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE006079; AAK02513.1; -
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001943; UVR.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF02151; UVR; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR SOS response; Excision nuclease; DNA repair; ATP-binding;
KW Complete proteome.
FT NP_BIND 44 51 ATP (POTENTIAL).
FT SEQUENCE 678 AA; 77401 MW; 2EE928B9A23ADF15 CRC64;

Query Match 4.2%; Score 89; DB 1; Length 678;
Best Local Similarity 21.1%; Pred. No. 14;
Matches 40; Conservative 38; Mismatches 82; Indels 30; Gaps 6;

Qy 9 RCPDYLLKNEY-----DDIVRSDVALDKQCK-IAQHPDVMLELQREKAAQMHLV 58
Db 401 PGP-----YELEKSGGBIIDQVVRPTGLDPEIRPVAIQVDDLLSEARQADQNERV 454
Qy 59 LKLEQFSNTYSNLTITPENNYRTMHGRAVNGSQLGDKYIQLKSLLOPIRYSRASLYGPN 118
Db 455 LV-----TTLTKKMAEDTLDLDEHGRVRYLHSDIDTVERVEIIRDLRGEFDVLVGIN 509
Qy 119 IGR-----PRKNVIALLD-----GFMKVAGSTVDVATWQHCHYIDGRVVKVMDFLKRLTLD 169
Db 510 LLREGLDIPEVSLVAILDADKEGFLRSERSLIQTIGRAARNLKGKAILYADRTINSQKA 569
Qy 170 LSDQIRKIQK 179
Db 570 IETNRREK 579
```

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:30:47 ; Search time 123.94 Seconds

(without alignments)  
565.297 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_130\_534

Perfect score: 2129

Sequence: 1 NPAKSRGGPGPDYLYKNYED.....PVTMGFFVKNVNALACRYR 405

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2129	100.0	534	4 Q9HB38	Q9hb38 homo sapien
2	2090	98.2	592	4 Q9HB37	Q9hb37 homo sapien
3	1822	85.6	480	4 Q9HB39	Q9hb39 homo sapien
4	810.5	38.1	536	11 Q9Q2F8	Q9qz2f8 rattus norv
5	795.5	37.4	543	4 Q9Y251	Q9y251 homo sapien
6	795.5	37.4	545	4 Q9UL39	Q9ul39 homo sapien
7	785	36.9	545	6 Q9MY0	Q9my0 bos taurus
8	736.5	34.6	523	13 Q9QYK5	Q9qyk5 gallus gall
9	301	14.1	521	10 Q9SDA1	Q9sda1 arabidopsis
10	301	14.1	543	10 Q9FF10	Q9ff10 arabidopsis
11	258.5	12.1	516	10 Q9FLK8	Q9flk8 arabidopsis
12	236	11.1	536	10 Q9FZP1	Q9fzp1 arabidopsis
13	215.5	10.1	527	10 Q9LRC8	Q9lrc8 scutellaria
14	155	7.3	190	10 Q82604	Q82604 arabidopsis
15	155	7.3	935	5 Q9VE79	Q9ve79 drosophila
16	110	5.2	174	10 Q9ATW5	Q9atw5 zea mays (m

17	103	4.8	768	3	Q9P8N9	Q9p8n9 cladosporiu
18	102.5	4.8	1260	5	O15639	O15639 dictyosteli
19	102	4.8	885	5	O61232	O61232 lymanaea sta
20	98	4.6	491	2	O08457	O08457 clostridium
21	98	4.6	1090	12	O91E95	O91e95 human rotav
22	97	4.6	1020	5	O25415	O25415 leishmania
23	97	4.6	4533	5	O9BIX3	O9bix3 tetrahymena
24	96.5	4.5	1429	2	O9LAP7	O9lap7 alteromonas
25	96	4.5	898	10	O9SIB9	O9sib9 arabidopsis
26	95.5	4.5	551	10	O38779	O38779 avena sativ
27	95.5	4.5	663	8	O9MJ69	O9mj69 physarum po
28	95.5	4.5	1281	11	O88915	O88915 mus musculus
29	95	4.5	456	5	O22152	O22152 caenorhabdi
30	95	4.5	512	8	O99806	O99806 ixodes hexa
31	95	4.5	557	16	O97G52	O97g52 clostridium
32	95	4.5	740	10	O82436	O82436 cucumis mel
33	94.5	4.4	460	16	O98QL8	O98ql8 mycoplasma
34	94	4.4	855	15	O90DZ7	O90dz7 human immun
35	93.5	4.4	1234	5	O9YIH7	O9yih7 dictyosteli
36	93.5	4.4	1219	11	O9QYM3	O9qym3 mus musculus
37	93.5	4.4	1713	9	O94MA1	O94ma1 lactococcus
38	93	4.4	320	17	O97UM5	O97um5 sulfolobus
39	93	4.4	441	10	O94I11	O94i11 arabidopsis
40	93	4.4	503	3	O9HGN1	O9hgn1 schizosacch
41	92.5	4.3	515	10	O49258	O49258 avena sativ
42	92.5	4.3	620	10	O9FUI4	O9fui4 oryza sativ
43	92.5	4.3	685	4	O96CV7	O96cv7 homo sapien
44	92.5	4.3	876	16	O99TH2	O99th2 staphylococ
45	92.5	4.3	1845	12	Q9WJZ8	Q9wjz8 human cytom

## ALIGNMENTS

RESULT 1

Q9HB38	PRELIMINARY;	PRT;	534 AA.
AC	Q9HB38;		
DT	01-MAR-2001 (Tremblrel. 16, Created)		
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)		
DT	01-MAR-2001 (Tremblrel. 16, Last annotation update)		
DE	HEPARANASE-LIKE PROTEIN HPA2B.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20483645; PubMed=11027606;		
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,		
RA	Hircock M., Patel S., Barry E., Stuberfield C., Terrett J., Page M.,		
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian		
RT	Heparanase Family Member."		
RL	Biochem. Biophys. Res. Commun. 276:1170-1177(2000).		
DR	EMBL; AF282886; AAG23422.1; -.		
SQ	SEQUENCE 534 AA; 60063 MW; C3DE5E90CB338C4 CRC64;		

Query Match 100.0%; Score 2129; DB 4; Length 534;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-174;  
 Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	NPAKSRGGPGPDYLYKNYEDIVRSVDALDKQKCKIAQHPDVMLEQREKAQMHVLVL 60
DB	130	NPAKSRGGPGPDYLYKNYEDIVRSVDALDKQKCKIAQHPDVMLEQREKAQMHVLVL 189
QY	61	KEQFSNTSYNLIITPENNYRTHGRAVNSQLGKDYQLKSLQPIRIYRSRSLYGNP 120
DB	190	KEQFSNTSYNLIITPENNYRTHGRAVNSQLGKDYQLKSLQPIRIYRSRSLYGNP 249
QY	121	RPKNVIALDGPVKVAGSTVDATVWQHCVIDGRVVKVMDFLKRLDLDTSQIRIKQV 180
DB	250	RPKNVIALDGPVKVAGSTVDATVWQHCVIDGRVVKVMDFLKRLDLDTSQIRIKQV 309

Query Match	98.2%	Score	2090;	DB	4;	Length	592;
Best Local Similarity	87.5%	Prod. No.	6.le-171;				
Matches	405;	Conservative	0;	Mismatches	0;	Indels	58;
Gaps							
Qy	1	NPAKSRGGPDPYILKNYEDD	IVRSDVALDKQCKIAQHDPVMLEQREAAQMHVLVLL	60			
Db	130	NPAKSRGGPDPYILKNYEDD	IVRSDVALDKQCKIAQHDPVMLEQREAAQMHVLVLL	189			
Qy	61	KQFSNTYSNLIUT	-----	74			
Db	190	KQFSNTYSNLIUT	ARSRLDKLYNFADCSGLHIFALNALRRPNNSWNSSALSLLKYS	249			
Qy	75	-----	EPNNYRTMHGRAVNGSOLGKDYIQLKSLLOPIRIYSRASLYGPNIGRP	122			
Db	250	SKKYNTSWEIGNEPNNYRTMHGRAVNGSOLGKDYIQLKSLLOPIRIYSRASLYGPNIGRP	309				
Qy	123	RKNVIALLDGFMKVAGSTVDVATVWQHCY	IDGRVVKVYMDFLKTRLLDITSDQIRKTKQVYN	182			
Db	310	RKNVIALLDGFMKVAGSTVDVATVWQHCY	IDGRVVKVYMDFLKTRLLDITSDQIRKTKQVYN	369			
Qy	183	TYTPGKKIWLLEGVYVTT	SAGGTNNLSDSYAAGFLWLNTLGLMLANOGIDVYIRHSFFDHGYN	242			
Db	370	TYTPGKKIWLLEGVYVTT	SAGGTNNLSDSYAAGFLWLNTLGLMLANOGIDVYIRHSFFDHGYN	429			
Qy	243	HLVDQNFNPDPYDWLSLLYKRLIGPKVLAVHVHAGVLQKRPGRVIRDKLRIYAHCTNNHN	302				
Db	430	HLVDQNFNPDPYDWLSLLYKRLIGPKVLAVHVHAGVLQKRPGRVIRDKLRIYAHCTNNHN	489				
Qy	303	HNVVRGSITLFIINLHRSRKKIKLACTLRDKLVHQVYLLLOPYGQEGLKSKSVOLNGOPLYM	362				
Db	490	HNVVRGSITLFIINLHRSRKKIKLACTLRDKLVHQVYLLLOPYGQEGLKSKSVOLNGOPLYM	549				

QY	363	VDDGTLPELKPRLRAGRTLVIIPVTMGFFVVKVNNALACRYR	405
Db	550	VDDGTLPELKPRLRAGRTLVIIPVTMGFFVVKVNNALACRYR	592
RESULT	3		
Q9HB39			
ID	Q9HB39	PRELIMINARY;	PRT; 480 AA.
AC	Q9HB39;		
DT	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)		
DE	HEPARANASE-LIKE PROTEIN HPA2A.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20483645; PubMed=11027606;		
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,		
RA	Hircock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;		
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian		
RT	HEPARANASE Family Member.";		
RL	Heparanase. Biochem. Res. Commun. 276:1170-1177(2000).		
DR	EMBL; AF28285; AAG23421.1;		
SQ	SEQUENCE 480 AA; 53900 MW; F75F8967AC1FF83 CRC64;		
Query Match	85.6%;	Score 1822;	DB 4; Length 480;
Best Local Similarity	86.7%;	Pred. No. 4.7e-148;	
Matches 351;	Conservative	0; Mismatches	0; Indels 54; Gaps
QY	1	NPAKSRGGPGPDYLLKNYEDDIVRSDVALDKQCKIAQHPDVMLELQREKAAQMHVLVL	60
Db	130	NPAKSRGGPGPDYLLKNYED	149
QY	61	KEQFSNTYSNLITPENRYRTHMGRAVNGSQLGKDYIQLKSLQPIRVSRASLYGPNIG	120
Db	150	-----EPNNRYTHMGRAVNGSQLGKDYIQLKSLQPIRVSRASLYGPNIG	195
QY	121	RPRKNVIALLDGFMKVGASTVDATWQHCVIDGRVVKVMDFLKTRLLDLSQIRKIQKV	180
Db	196	RPRKNVIALLDGFMKVGASTVDATWQHCVIDGRVVKVMDFLKTRLLDLSQIRKIQKV	255
QY	181	VNTYTPGKKIWLEGVVTTTSAGCTNNLSDSYAAGFLWNTGLMNAQIDVIRHSFFDHG	240
Db	256	VNTYTPGKKIWLEGVVTTTSAGCTNNLSDSYAAGFLWNTGLMNAQIDVIRHSFFDHG	315
QY	241	YNHLVDQNFPLDPYWLSSLYKELIGPKVLAVHVAGLQRPGRPRVIRDKLRIYHCTNH	300
Db	316	YNHLVDQNFPLDPYWLSSLYKELIGPKVLAVHVAGLQRPGRPRVIRDKLRIYHCTNH	375
QY	301	HNHNVRGSIITFLINLHSRKKIKLAGTLRDKLVHQLQPYQGBELKSKSVQLNGQPL	360
Db	376	HNHNVRGSIITFLINLHSRKKIKLAGTLRDKLVHQLQPYQGBELKSKSVQLNGQPL	435
QY	361	VWVDDGTLPELKPRLRAGRTLVIIPVTMGFFVVKVNNALACRYR	405
Db	436	VWVDDGTLPELKPRLRAGRTLVIIPVTMGFFVVKVNNALACRYR	480
RESULT	4		
Q9QZF8			
ID	Q9QZF8	PRELIMINARY;	PRT; 536 AA.
AC	Q9QZF8;		
DT	01-WAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	HEPARANASE.		
OS	Rattus norvegicus (Rat).		
GN	Rattus norvegicus (Rat).		

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN NCBI\_TaxID=10116;  
RP SEQUENCE FROM N.A.  
RA Podyma K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;  
RT "Heparanase from parathyroid cell line."  
RL Submitted (Sep-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF184967; AAF04563.1; -  
SQ SEQUENCE 536 AA; 60569 MW; 6208B1FD9EE28421 CRC64;

Query Match 38.18; Score 810.5; DB 11; Length 536;  
Best Local Similarity 40.68; Pred. No. 4.6e-61;  
Matches 170; Conservative 66; Mismatches 114; Indels 69; Gaps 6;

Qy 42 DVMLELOREKAAQHLVLLKQF-----SNTYS-----NLIL- 73  
Db 128 DVLRKLQWEPFQ-ELLLRQYQREFKNSYSSVDMLYSFAKCSRLDLIFGLNALLR 186  
Qy 74 -----TEPNNRTMHGRAVNGSOLGKDYIQLKSL 103  
Db 187 TPDLRWNSNAQLLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQLGDFVELHKL 246

Qy 104 QPIRIYRASLYGNIGPRKNTVALLDGFPMKVGAGSTVDATWQHCYIDGRVVKVMDFLK 163  
Db 247 QK-SAFONAKLYGPDIGQPRGKTWKLLRSFLKAGGEVIDSLTWHRYLNGRVATKREDFLS 305

Qy 164 TRLDLTLSQIRKIOKVNTVTPGKKIWLEGVVTTSAGTNNLSYAAAGFLWLTGLMQL 223  
Db 306 SDVLDTFILSVQKILKVTKEPTGKVKWLGTSAYGGAPLLSTFAAGFMWLDKLGLS 365

Qy 224 ANQIDVVIHRSFFDHGYNHLVDQNFNPDYWLSSLKRLIGPKVLAHVAGLQKRP 283  
Db 366 AOLGIEVVMQVFEGAGNYHLVDENFELPDYWLSSLKRLIGPKVLAHVAGLQKRP 420

Qy 284 GRVTRDKLRIYAHCTNNHNNHNVGSGITFLINLHRSRKKTKLACTLRDKLVHQLLOPY 343  
Db 421 ----RSKLRYLVHCTNVYHPYREGDLFLYVLNLHNVTKHLKLPPEFSRPVDKLYLLKPF 476

Qy 344 QOGLKSKSVOLNGOPLVMVDGTLPELKPRLPRAGRTLVIPTVMTGFFVKNVNALAC 402  
Db 477 GSDGLKSVOLNGOPLVMVDGTLPELKPRLPRAGRTLVIPTVMTGFFVKNVNALAC 535

RESULT 5  
ID QY251 PRELIMINARY; PRT; 543 AA.  
AC QY251;  
DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)  
DE HEPARANASE.  
GN HPA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OX Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN NCBI\_TaxID=9606;  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RA Hulet M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,  
RA Parish C.R.;  
RT "Cloning of mammalian heparanase, an important enzyme in tumor  
RT invasion and metastasis."  
RL Nat. Med. 5:803-809(1999).  
RP SEQUENCE FROM N.A.  
RA Vlodavsky I., Friedman Y., Elkin M., Aingorn H., Atzmon R.,  
RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,  
RA Spector L., Pecker I.;  
RT "Mammalian heparanase: a novel gene involved in tumor progression and

RT metastasis.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN NCBI\_TaxID=10116;  
RP SEQUENCE FROM N.A.  
RA MEDLINE=993377052; PubMed=10446189;  
RA Toyoshima M., Nakajima M.;  
RT "Human heparanase. Purification, characterization, cloning, and  
RT expression."  
RL J. Biol. Chem. 274:24153-24160(1999).  
RN NCBI\_TaxID=9606;  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RA MEDLINE=99333379; PubMed=10405343;  
RA Kussie P.H., Hulmes J.D., Ludwig D., Patel S., Navarro E.C.,  
RA Seddon A.P., Giorgio N.A., Bohlen P.;  
RT "Cloning and Functional Expression of a Human Heparanase Gene."  
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).  
DR EMBL; AF165154; AAD45379.1; -  
DR EMBL; AF144325; AAD41342.1; -  
DR EMBL; AF155510; AAD45494.1; -  
DR EMBL; AF152376; AAD45669.1; -  
SQ SEQUENCE 543 AA; 61176 MW; AD262EC267334AB2 CRC64;

Query Match 37.4%; Score 795.5; DB 4; Length 543;  
Best Local Similarity 37.8%; Pred. No. 9e-60;  
Matches 179; Conservative 66; Mismatches 139; Indels 89; Gaps 10;

Qy 1 NPAKSR-GPGPDYVL-----KNYEDDIYRSVDALDKQCKTAQ-HPDVMLEL 47  
Db 88 SPAYLRFGGTKTDFLIFDPKKESTFEERSYQSQVNQDI-----CKYGSIPDPVEKL 140

Qy 48 OREKAAQHLVLLKQF-----SNTYS-----NLIL----- 73  
Db 141 RLEWYQQL-LRREHYQKFKNSYSSSDVLYTFANCGLDLIFGLNALRLTADLOW 199

Qy 74 -----TEPNNRTMHGRAVNGSOLGKDYIQLKSLLOPIRY 109  
Db 200 NSSNAQLLLDYCSSKGYNISWELGNEPNSFLKADIFINGSOLGDFIQLHKLK- 258

Qy 110 SRASLYGNIGPRKNTVALLDGFPMKVGAGSTVDATWQHCYIDGRVVKVMDFLKTLDT 169  
Db 259 KNAKLYGPDVGPGRKRTAKMLKSLKAGGEVIDSVTWHRYLNGRTATREDLNPDLVDI 318

Qy 170 LSDQIRKIOKVNTVTPGKKIWLEGVVTTSAGTNNLSYAAAGFLWLTGLMLANQGIT 229  
Db 319 FTSSVQKVFQVVESTRPGKKVWLGTSAYGGAPLLSDTFAAGFMWLDKLGSLRMGLE 378

Qy 230 VVIRHSFFDHGYNHLVDQNFNPDYWLSSLKRLIGPKVLAHVAGLQKRPGRVIRD 289  
Db 379 VVMQVFFGAGNYHLVDENFDPDYWLSSLKRLIGPKVLAHVAGLQKRPGRVIRD 429

Qy 290 KLRIYAHCTNNHNNHNVGSGITFLINLHRSRKKTKLACTLRDKLVHQLLOPYGOEGLK 349  
Db 430 KURVYLHCTNTDNPYKREGDLFLYVLNLHNVTKHLKLPPEFSRPVDKLYLLKPF 489

Qy 350 SKSVOLNGOPLVMVDGTLPELKPRLPRAGRTLVIPTVMTGFFVKNVNALAC 402  
Db 490 SKSVOLNGOPLVMVDGTLPELKPRLPRAGRTLVIPTVMTGFFVKNVNALAC 542

RESULT 6  
ID Q9UL39 PRELIMINARY; PRT; 545 AA.  
AC Q9UL39;  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE HEPARANASE.  
GN HPA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OX Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN NCBI\_TaxID=9606;

```

RN  SEQUENCE FROM N.A.
RP  TISSUE=PLACENTA;
RC  MEDLINE=20229546; PubMed=10764835;
RX  Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;
RA  "Heparanase expression in invasive trophoblasts and acute vascular
RT  damage.";
RL  Glycobiology 10:467-475 (2000).
DR  EMBL; AF084467; AAB54516.1; -.
SQ  SEQUENCE 545 AA; 61418 MW; 67B80ACD73C5A9A1 CRC64;

Query Match 37.48; Score 795.5; DB 4; Length 545;
Best Local Similarity 37.88; Pred. No. 9.1e-60;
Matches 179; Conservative 66; Mismatches 139; Indels 89; Gaps 10;

QY 1 NPAKSR-GSPGPDYVL-----KNYEDDIVRSVALDKQKCKIAQ-HPDVMLEL 47
DB 90 SAYLRFEGTKTDFLIFDPKKESTFEERSYQSQVNQDI-----CKYGSIPDPVEKL 142
QY 48 QREKAAQHLVLLKEQF-----SNTYS-----NLIL----- 73
DB 143 RLEWYPQEQ-LLREHYQKFKNSTYSRSDVLYTFANCGLDLIFGLNALLRTADLQW 201
QY 74 -----TEPNNYRTMHGRAVNGSQLGKDYIQLKSLQPIRIY 109
DB 202 NSSNAQLLDYCSSKGYNISWELGNEPNSFLLKADIFINGSQLGDFIQLHKLK-SPF 260
QY 110 SRASLYGPNIGRPNKVNIALDGFMKVAGSTVDVATVQHCYIDGRVVKVMDFLKTRLLDT 169
DB 261 KNKLYGPDVGQPRKTKAKMLKSLKAGEVIDSVTWHYVNGRTATREDFLNPDVLDI 320
QY 170 LSDQIRKIQVNTYTPGKKIWEVVTTSAGGTNNLSDSYAAGFLWNTLGLMLANQIGD 229
DB 321 FISSVQKVFQVVESTRPGKVMWLGTSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIE 380
QY 230 VVTRHSFEFHGYHLVDQNFNPLPDVWLSLLYKRLIGPKVLAVHVGALQKRPGRVIRD 289
DB 381 VNRQVFFGAGNVHLVDENFDLPDWLSLLFKLGVTKVLMASVQSGKRR----- 431
QY 290 KLRIYAHCTNNHNNVRSITFLFIINLHRSRKKIKLAGTLRDLKVLQVHQLLPYQOGLK 349
DB 432 KLRYLHCTNTDNPYKEGDLTLYALNLHNVTKYLRPLPYFSNKKQVQDKYLLRPLGPHGLL 491
QY 350 SKSVQLNGQLVWVDGTLPELKPRLAGRTLVIPPTMTGFFVVKVNNALAC 402
DB 492 SKSVQLNGQLTKMVDQDTLPPLMEKPLRPGSSLGIPAFSYFFVIRNAKVAAC 544

RESULT 7
ID Q9MYX0 PRELIMINARY; PRT; 545 AA.
AC Q9MYX0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE HEPARANASE.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Kizaki K., Nakano H., Takahashi T., Imai K., Hashizume K.;
RT "Expression of Heparanase mRNA in Bovine Placenta During Gestation.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ Databases.
DR EMBL; AF281160; AAB7301.2; -.
SQ SEQUENCE 545 AA; 61076 MW; FAC4BDFD855B933 CRC64;

Query Match 36.98; Score 785; DB 6; Length 545;

```

```

Best Local Similarity 37.0%; Pred. No. 7.3e-59;
Matches 170; Conservative 65; Mismatches 145; Indels 80; Gaps 6;

QY 1 NPAKSRGGPGPDYVLKYNEDDIVRSVALDKQKCKIAQHPDVMLELOREKAAQHLVLL 60
DB 107 DPKEPAFEERSYWLQSQNODI-----CKSGSIPSDVEEKLRLWPFQFOVLL 154
QY 61 KEQF-----SNTYS-----NLIL----- 73
DB 155 REQYQKKFTNSTYSRSDVLYTFASCGLNLIFGVNALLRTTDMHWDSSNAQLLDYCS 214
QY 74 -----TEPNNYRTMHGRAVNGSQLGKDYIQLKSLQPIRIYSRASLYGPNIGRP 122
DB 215 SKYVNISWELGNEPNSFORKAGIFINGRQLGDEFIEFRKLLGK-SAFKNKAKLYGPDIGOP 273
QY 123 RKNVIALDGFMKVAGSTVDVATVQHCYIDGRVVKVMDFLKTRLLDTLSQIRKIQKVVN 182
DB 274 RRTVTKMLKSLKAGEVIDSVTWHYVNGRTATREDFLNPDILDTFISVQKTLRIVE 333
QY 183 TYTPGKKIWEVVTTSAGGTNNLSDSYAAGFLWNTLGLMLANQIGDVIIRHSFEFHGYN 242
DB 334 KIRPLKVKWLGTSAYGGGAPLLSNTFAAGFMWLDKGLSARMGIEVVMRQVLFAGNY 393
QY 243 HLVDQNFNPLPDVWLSLLYKRLIGPKVLAVHVGALQKRPGRVIRDKLRIYAHCTNNH 302
DB 394 HLVDGNEPLPDVWLSLLFKKLVGNKVLMSVKGPD-----RSKFRVYLHCTNTKH 444
QY 303 HNYVRSITFLFIINLHRSRKKIKLAGTLRDLKVLHVOYLLQPVGOBGLSKSVQLNGQLM 362
DB 445 PRYREGDITLVALNLHNVTKHLELPHLFFNKQVDKYLKPSGTDGLLSKSVQLNGQLM 504
QY 363 VDDGTLPKLPRLAGRTLVIPPTMTGFFVVKVNNALAC 402
DB 505 VDEQTLPALTEKPLHPSGLMPPPSYGFYFVIRNAKVAAC 544

RESULT 8
ID Q90YK5 PRELIMINARY; PRT; 523 AA.
AC Q90YK5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE HEPARANASE.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11387326;
RA Goldshmidt O., Zcharia E., Alingorn H., Guatta-Rangini Z., Atzmon R.,
RA Michal I., Pecker I., Mitrani E., Vlodavsky I.;
RT "Expression Pattern and Secretion of Human and Chicken Heparanase Are
RT Determined by Their Signal Peptide Sequence.";
RL J. Biol. Chem. 276:29178-29187 (2001).
DR EMBL; AY037007; AAK82648.1; -.
SQ SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF881 CRC64;

Query Match 34.6%; Score 736.5; DB 13; Length 523;
Best Local Similarity 44.2%; Pred. No. 1e-54;
Matches 145; Conservative 53; Mismatches 121; Indels 9; Gaps 2;

QY 75 EPNNYRTMHGRAVNGSQLGKDYIQLKSLQPIRIYSRASLYGPNIGRPKNVIALDGF 134
DB 204 EPNSFKKSGICIDGFGQLGRDFVHLRQLLSQHPLYRHAELYGLDVGQPKRHTQHLLRSFM 263
QY 135 KVAGSTVDVATVQHCYIDGRVVKVMDFLKTRLLDTLSQIRKIQKVVNTYTPGKKIWE 194
DB 264 KSGGAIDSVTWHYVNGRSATREDFLSPEDLSFATAIHDLVGLIVEATVPFGKKWLG 323

```



QY 195 VVTSAGTNNLSDSYAAGFLWNTLGLMANGIDVIRHSEFFDHGYNHLVDQNFPLPD 254  
Db 324 TGSAYGGAPQLSNTYVAGFMWLDKLGAAARRGIDVVMRQVSFGAGSYHLVDAGFKPLPD 383  
QY 255 YWLSLLYKRLGPKVLAVHVGAGLQKPPGRVIRDKLRIYAHCTNHHNHNXVSGSITLFI 314  
Db 384 YWLSLLYKRLVTRVLQNSVEQADAR-RP-----RVLHCTNPRPKKREGDVTILFA 434  
QY 315 INLHRSKKIKLAGTLRDLKLVHQLYQGLKSKSVQNLGQPLVMVDGTLPELKPR 374  
Db 435 LNLNSVTSQSLQPKOLWSKSDVYLLPHGKDSILSREVQLNGRLQLQVDDTLPALHEM 494  
QY 375 PLRAGRILVIPPVTMGFFVKNVNALAC 402  
Db 495 ALAPGSGTLGLPAFSYGFYVIRNAKAIAC 522

RESULT 9  
Q9SDA1 PRELIMINARY; PRT; 521 AA.

AC Q9SDA1;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE HYPOTHETICAL 57.8 KDA PROTEIN.  
GN F13G24.30.

OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Van Der Schueren J., Chuang Y.J., Voet M., Robben J.,  
RA Volckaert G., Bancroft I., Mewes H.W., Lemcke K., Mayer K.F.X.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL133421; CAB62595.1; -.  
DR InterPro; IPR001254; Trypsin.  
DR PROSITE; PS00135; TRYPSIN\_SER; UNKNOWN\_1.  
KW Hypothetical protein.  
SQ SEQUENCE 521 AA; 57831 MW; 07D8664A4B305CC2 CRC64;

Query Match 14.1%; Score 301; DB 10; Length 521;  
Best Local Similarity 28.2%; Pred. No. 2.6e-17;  
Matches 105; Conservative 50; Mismatches 127; Indels 90; Gaps 15;

QY 84 GRAVNGSQLGRDYIOLKSLLOPIRIYSRASLYGPNIGRPNKVNIALLDGF-----M 134  
Db 186 GASVSAELYGKDLVLKDVIN--KVYKNSWLHKPLVAP-----GGFYEQQWTKLL 235  
QY 135 KVAG-STVDVAVTWHQCYIDGR-----VVKVMDFLKTLRLDLSQIRKIQKVVNTYTPG 187  
Db 236 EISGPSVVDVVT-HHIYNLGSGNDPALVKKIMD---PSYLSQVSKTFKDVNQTIOEHGP- 290  
QY 188 KKIWLEGVVTTSAG-----GTNNLSDSYAAGFLWNTLGLMANGIDVIRHSEFFDHGYNH 243  
Db 291 ---RAGVTLLNLNSQSDFTVSNSGINVVLNAESRKKKSLDTLKRPFSWIGSKASDG 450  
QY 244 LVDQNFNPLDPYWLKSLYKRLIGPKVLAVHVGAGLQKPPGRVIRDKLRIYAHCTNHHNH 303  
Db 348 LEKGTFFVNPDPYISALLHRLMGKGLAVQTDG-----PPQLRVYAHCSKG--- 393  
QY 304 NVYRGSTIFLFIINH-----RSRKKIKLAGTLRDLKLV----- 335  
Db 394 ---RAGVTLLNLNSQSDFTVSNSGINVVLNAESRKKKSLDTLKRPFSWIGSKASDG 450  
QY 336 ---HOYLLQPVGQEG-LKSKSVQLNGQPLVMVDGTLPELKPRPLRAGRITLVIPTVMG 390  
Db 390 ---HOYLLQPVGQEG-LKSKSVQLNGQPLVMVDGTLPELKPRPLRAGRITLVIPTVMG 390

Db 451 YLNREYHLTP--ENGVLRSKTMVLNGSKLSKPTATGDIPLSLEPVLRSVNSPLNVLPLSMS 508  
QY 391 FFVKNVNALAC 402  
Db 509 FIVLPNFDASAC 520

RESULT 10  
Q9FF10 PRELIMINARY; PRT; 543 AA.

AC Q9FF10;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE SIMILARITY TO HEPARANASE.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=COLUMBIA;  
RX MEDLINE=97471969; PubMed=9330910;  
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,  
RA Miyajima N., Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence  
RT features of the 1.6 Mb regions covered by twenty physically assigned  
RT pl clones";  
RL DNA Res. 4:215-230(1997).  
DR EMBL; AB005249; BAB09947.1; -.  
DR InterPro; IPR001254; Trypsin.  
DR PROSITE; PS00135; TRYPSIN\_SER; UNKNOWN\_1.  
SQ SEQUENCE 543 AA; 60250 MW; 0FA2248948282FF6 CRC64;

Query Match 14.1%; Score 301; DB 10; Length 543;  
Best Local Similarity 28.2%; Pred. No. 2.8e-17;  
Matches 105; Conservative 50; Mismatches 127; Indels 90; Gaps 15;

QY 84 GRAVNGSQLGRDYIOLKSLLOPIRIYSRASLYGPNIGRPNKVNIALLDGF-----M 134  
Db 208 GASVSAELYGKDLVLKDVIN--KVYKNSWLHKPLVAP-----GGFYEQQWTKLL 257  
QY 135 KVAG-STVDVAVTWHQCYIDGR-----VVKVMDFLKTLRLDLSQIRKIQKVVNTYTPG 187  
Db 258 EISGPSVVDVVT-HHIYNLGSGNDPALVKKIMD---PSYLSQVSKTFKDVNQTIOEHGP- 312  
QY 188 KKIWLEGVVTTSAG-----GTNNLSDSYAAGFLWNTLGLMANGIDVIRHSEFFDHGYNH 243  
Db 313 ---WASPVWGESGGAYNSGGRHVSOTFIDSFYLDQLGMSARHNTKVYCRQTLVGGFYGL 369  
QY 244 LVDQNFNPLDPYWLKSLYKRLIGPKVLAVHVGAGLQKPPGRVIRDKLRIYAHCTNHHNH 303  
Db 370 LEKGTFFVNPDPYISALLHRLMGKGLAVQTDG-----PPQLRVYAHCSKG--- 415  
QY 304 NVYRGSTIFLFIINH-----RSRKKIKLAGTLRDLKLV----- 335  
Db 416 ---RAGVTLLNLNSQSDFTVSNSGINVVLNAESRKKKSLDTLKRPFSWIGSKASDG 472  
QY 336 ---HOYLLQPVGQEG-LKSKSVQLNGQPLVMVDGTLPELKPRPLRAGRITLVIPTVMG 390  
Db 473 YLNREYHLTP--ENGVLRSKTMVLNGSKLSKPTATGDIPLSLEPVLRSVNSPLNVLPLSMS 530

QY 391 FFVKNVNALAC 402  
Db 531 FIVLPNFDASAC 542

RESULT 11  
Q9FLK8 PRELIMINARY; PRT; 516 AA.

ID Q9FLK8  
AC Q9FLK8;



```

Query Match      7.3%; Score 155; DB 5; Length 935;
Best Local Similarity 22.6%; Pred. No. 0.00022;
Matches 84; Conservative 54; Mismatches 130; Indels 104; Gaps 21;

Qy 76 PNNYRTWHG-RAVNGS-----QLG-----KDYIQ-----LKSLQPIRIY-SR 111
      | : : : | : | | | : | : | : |
Db 85 PDDWDSMHTLKILNTSYMGVITDCIWLQGLDFGTSRADYVQELRTLKLTMDTFKPYDD 144

```

```
QY 112 ASLYGPNIGRPNKVNIALLDGFMKVAGSTVDATWQHCVID-----GRVVKVMDFL 162
Db 145 WRLMGADIS-----AGSSADET---KRYVDMKDLNTAFGWTQTPANMLP 185
QY 163 KTRLLDTLSDQIRKIQKVVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWLNTLGM 222
Db 186 KSSLSGLYSDPALRTLQQQRVP---LWL--TLPEERSQRLVGDETTDALRWAQTMGD 240
QY 223 LANQGDVVRHSFFDHGYNHLVD---QNFNPLPDYWLSLLYKRLIGPKVLAVHVAGLQR 279
Db 241 AAASGFDVIFKRM-----NLVDFERPNS----LYVTALFKVVGSRVFFA----- 282
QY 280 KPRPGRVIRDKLRIYAHCTNHHNHVYRGSITLFIINLHRSRKKI--KLAGTLRDKLVHQ 337
Db 283 --RPLNAFAPSNNKLYTHCA-----NAVSGGLAFMVVNTTEEQPTTIVKSTSLSSSEIWQ 335
QY 338 YLQPYGQEGLSKSKSVQLNGQPLVMYDDCTL-PELKP-----RPLRAGRPLVIPPTMGFF 392
Db 336 YVLTGHDQR-----VOLNNVRLHL--NTTLRPLIKPIDPTKPLQ----LITPSMAVSEFW 383
QY 393 VVKNNVALACRY 404
Db 384 VLPDVNLEHCQF 395
```

Search completed: July 30, 2002, 08:30:49  
Job time: 1001 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:15:05 ; Search time 53.59 Seconds  
(without alignments)  
184.593 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_130\_534

Perfect score: 2129  
Sequence: 1 NPAKSRGGPGPDYILKNYED.....PVTMGFFVYKYNALACRYR 405

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	805	37.8	380	4	US-09-181-336-19
2	799.5	37.6	543	2	US-08-922-170B-10
3	799.5	37.6	543	4	US-09-071-739B-2
4	799.5	37.6	543	4	US-09-260-038B-2
5	795.5	37.4	543	4	US-09-181-336-13
6	779	36.6	380	4	US-09-181-336-17
7	752.5	35.3	532	4	US-09-181-336-15
8	93.5	4.4	934	1	US-08-215-805A-80
9	92.5	4.3	685	2	US-08-878-989-1
10	92.5	4.3	685	3	US-08-136-282-2
11	92.5	4.3	685	4	US-09-272-796-1
12	92.5	4.3	685	4	US-09-505-744-2
13	88	4.1	1106	1	US-08-435-675B-5
14	88	4.1	1106	1	US-08-336-257A-8
15	86	4.0	1103	1	US-08-455-543A-53
16	86	4.0	1103	2	US-08-223-305C-53
17	85	4.0	673	4	US-09-078-347A-2
18	84.5	4.0	434	2	US-09-989-925-1
19	84	3.9	538	4	US-09-175-928-4
20	83	3.9	490	4	US-08-740-223A-12
21	83	3.9	497	4	US-08-740-223A-4
22	83	3.9	834	1	US-08-471-033-21
23	83	3.9	834	2	US-08-471-044-21
24	83	3.9	834	2	US-08-463-483A-21
25	83	3.9	834	2	US-08-471-046A-21
26	83	3.9	834	2	US-08-470-566B-21
27	83	3.9	834	2	US-08-469-334-21

28	83	3.9	834	3	US-09-300-529-21	Sequence 21, Appl
29	82.5	3.9	921	1	US-07-872-644-39	Sequence 39, Appl
30	82.5	3.9	921	1	US-08-297-494-39	Sequence 39, Appl
31	82.5	3.9	921	1	US-08-297-510-39	Sequence 39, Appl
32	82.5	3.9	921	1	US-08-479-532-39	Sequence 39, Appl
33	82.5	3.9	921	1	US-08-455-526-39	Sequence 39, Appl
34	82.5	3.9	921	1	US-08-455-525-39	Sequence 39, Appl
35	82.5	3.9	921	3	US-09-139-491-39	Sequence 39, Appl
36	82.5	3.9	921	5	PCT-US92-03222-39	Sequence 39, Appl
37	82.5	3.9	942	1	US-07-872-644-43	Sequence 43, Appl
38	82.5	3.9	942	1	US-08-297-494-43	Sequence 43, Appl
39	82.5	3.9	942	1	US-08-297-510-43	Sequence 43, Appl
40	82.5	3.9	942	1	US-08-479-532-43	Sequence 43, Appl
41	82.5	3.9	942	1	US-08-455-526-43	Sequence 43, Appl
42	82.5	3.9	942	1	US-08-455-525-43	Sequence 43, Appl
43	82.5	3.9	942	3	US-09-139-491-43	Sequence 43, Appl
44	82.5	3.9	942	5	PCT-US92-03222-43	Sequence 43, Appl
45	82	3.9	497	1	US-08-373-579-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1  
US-09-181-336-19  
; Sequence 19, Application US/09181336A  
; Patent No. 6242238  
; GENERAL INFORMATION:  
; APPLICANT: FREEMAN, Craig Geoffrey  
; APPLICANT: PARISH, Richard  
; APPLICANT: HAMDORF, Brenton James  
; APPLICANT: HULETT, Mark Darren  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN  
; TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR  
; FILE REFERENCE: 032505-004  
; CURRENT APPLICATION NUMBER: US/09/181.336A  
; CURRENT FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: AU PP0062  
; EARLIER FILING DATE: 1997-10-28  
; EARLIER APPLICATION NUMBER: AU PP0812  
; EARLIER FILING DATE: 1997-12-09  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Rattus sp.  
US-09-181-336-19

Query Match	37.8%	Score 805;	DB 4;	Length 380;
Best Local Similarity	47.3%	Pred. No. 8.5e-79;		
Matches 155;	Conservative 57;	Mismatches 106;	Indels 10;	Gaps 2;
QY	75	EPNVTWTHGRAVNSQGLKDYIQLKSLLOPTIRYSRASLYGPNIGRPNKVNIALLDGFM	134	
DB	62	EPNFWKKAHISIDGLQGEDEFELKLIQK-SAFQNAKLYGPDIGQPRGKTVKLLRSPL	120	
QY	135	KVAGSTVDATVWHQCYIDGRVVKVMDFLKTLRLDLSQIRKIQKVNTYTPGKKIWLBG	194	
DB	121	KAGEVIDSLTWHHYLLNGRVATKEDFLSSVDLDTFILSVQKILVKTKEMTPGKKVWLGE	180	
QY	195	VVTSAGGTNNLSDSYAAGFLMLNTLGMLANOGIDVIRHSFFDGHYNHLVDQNFNPLPD	254	
DB	181	TSSAYCGGAPLSDTFAAGFMWLDKGLSAQLGIEVWVRQVFFGAGNYHLVDENEPLPD	240	
QY	255	YWLSTLYKRLIGPKVLAVHAGLRKPRGRVIRDKLRIYAHCTNHHNNHYVGSITLFI	314	
DB	241	YWLSTLLFKLVGPKVLMRSRVKGPD-----RSKLRVYLHCTNVYHPRYREGDLTYV	291	
QY	315	INLHRSRKKIKLAGLRDKLVHQYLLQPVQGEGLSKSVQLNGOPLVWVDDSTLPELKR	374	
DB	292	LNLHNVTKHLKLPMPFSPVDKYLKPKFGSDGLLSKSVQLNGOTLKMVDEQTLPALTEK	351	





US-09-181-336-13									
Query Match 37.4%; Score 795.5; DB 4; Length 543;									
Best Local Similarity 37.8%; Pred. No. 1.6e-77;									
Matches 179; Conservative 66; Mismatches 139; Indels 89; Gaps 10;									
QY	1	NPAKSR-GGPGDYYL-----KNYEDDIVRSVDVALDKGCKIAQ-HPDVMLEL	47						
Db	88	SPAYLRFGGTKTDFLFDPKKXESTFEERSYQSQVNQDI-----CKYGSIPDPVEEKL	140						
QY	48	QREKAAQMHVLVLLKEQF-----SNTYS-----NLIL-----	73						
Db	141	RLWPYQEQ-LLRHYQKKFNKSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLOW	199						
QY	74	-----TEPNYRTMHGRAVNSQLGKDYIQLKSLLOPIRY	109						
Db	200	NSSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFINGSQLGEDFIQLHLLRK-STF	258						
QY	110	SRLSYGPNIGRPRKNVIALLDGFMKRVAGSTVDVATWQHICYIDGRVVKVMDFLKTRLLDT	169						
Db	259	KNAKLYGPDVGQPRRTAKMLKSFLKAGGEVDSVTWHYLYNGRTATREDFLNPDVLDI	318						
QY	170	LSQDTRKIOKVNTYTPGKKIWLGVVTT'SAGTNNLSDSYAAGFLWLTGLMLANQGD	229						
Db	319	FISVQKVFQVVESTRPGKKVNLGETSSAYGGAPLLSDTFAAGFMWLDKGLSARMGIE	378						
US-09-181-336-17									
Query Match 36.6%; Score 779; DB 4; Length 380;									
Best Local Similarity 46.3%; Pred. No. 5.6e-76;									
Matches 152; Conservative 56; Mismatches 110; Indels 10; Gaps 2;									
QY	75	EPNNYRTMHGRAVNSQLGKDYIQLKSLLOPIRYSRLSYGPNIGRPRKNVIALLDGFM	134						
US-09-181-336-15									
Query Match 35.3%; Score 752.5; DB 4; Length 532;									
Best Local Similarity 38.0%; Pred. No. 7.2e-73;									
Matches 172; Conservative 62; Mismatches 130; Indels 89; Gaps 10;									
QY	1	NPAKSR-GGPGDYYL-----KNYEDDIVRSVDVALDKGCKIAQ-HPDVMLEL	47						
Db	88	SPAYLRFGGTKTDFLFDPKKXESTFEERSYQSQVNQDI-----CKYGSIPDPVEEKL	140						
QY	48	QREKAAQMHVLVLLKEQF-----SNTYS-----NLIL-----	73						
Db	141	RLWPYQEQ-LLRHYQKKFNKSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLOW	199						
QY	74	-----TEPNYRTMHGRAVNSQLGKDYIQLKSLLOPIRY	109						
Db	200	NSSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFINGSQLGEDFIQLHLLRK-STF	258						
QY	110	SRLSYGPNIGRPRKNVIALLDGFMKRVAGSTVDVATWQHICYIDGRVVKVMDFLKTRLLDT	169						
Db	259	KNAKLYGPDVGQPRRTAKMLKSFLKAGGEVDSVTWHYLYNGRTATREDFLNPDVLDI	318						
QY	170	LSQDTRKIOKVNTYTPGKKIWLGVVTT'SAGTNNLSDSYAAGFLWLTGLMLANQGD	229						
Db	319	FISVQKVFQVVESTRPGKKVNLGETSSAYGGAPLLSDTFAAGFMWLDKGLSARMGIE	378						
US-09-181-336-15									
Query Match 35.3%; Score 752.5; DB 4; Length 532;									
Best Local Similarity 38.0%; Pred. No. 7.2e-73;									
Matches 172; Conservative 62; Mismatches 130; Indels 89; Gaps 10;									
QY	1	NPAKSR-GGPGDYYL-----KNYEDDIVRSVDVALDKGCKIAQ-HPDVMLEL	47						
Db	88	SPAYLRFGGTKTDFLFDPKKXESTFEERSYQSQVNQDI-----CKYGSIPDPVEEKL	140						
QY	48	QREKAAQMHVLVLLKEQF-----SNTYS-----NLIL-----	73						
Db	141	RLWPYQEQ-LLRHYQKKFNKSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLOW	199						
QY	74	-----TEPNYRTMHGRAVNSQLGKDYIQLKSLLOPIRY	109						
Db	200	NSSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFINGSQLGEDFIQLHLLRK-STF	258						
QY	110	SRLSYGPNIGRPRKNVIALLDGFMKRVAGSTVDVATWQHICYIDGRVVKVMDFLKTRLLDT	169						
Db	259	KNAKLYGPDVGQPRRTAKMLKSFLKAGGEVDSVTWHYLYNGRTATREDFLNPDVLDI	318						
QY	170	LSQDTRKIOKVNTYTPGKKIWLGVVTT'SAGTNNLSDSYAAGFLWLTGLMLANQGD	229						
Db	319	FISVQKVFQVVESTRPGKKVNLGETSSAYGGAPLLSDTFAAGFMWLDKGLSARMGIE	378						







SEQUENCE CHARACTERISTICS:  
LENGTH: 685 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: HUVENB01  
CLONE: 39043  
US-09-272-796-1

Query Match 4.3%; Score 92.5; DB 4; Length 685;  
Best Local Similarity 23.2%; Pred. No. 0.43; 139; Indels 89; Gaps 18;  
Matches 83; Conservative 46; Mismatches 139; Indels 89; Gaps 18;

Qy 6 RGGPGPDYILKNYEDDIVRSVDALDKQKCKIAQHPDVMLELQREKAAQ-----M 55  
Db 90 KGG-----FAKCYE---MTDLTNKVVAAKIIPIHSRVAKPHQREKIDKEIHLRHLHK 140  
Qy 56 HLVLKKEQFSNTYNNLIITPEPNNTYTHGRAVNGSOLGKDYIQLKSLQIP-IRIYSRASL 114  
Db 141 HVQFYHYFEDKENIYILLEYCSRSM-----AHILK---ARKVLTEPEVRYLQIV 190  
Qy 115 YGPNIGRPRKNVIALLDGFMKVGASTVDVATWQHCYIDGRVVKVMDFLKTRLLDLSQI 174  
Db 191 SGLKYLHEQE---ILHRDLKLGNNFFIN-----EAMELKVGDFGLAARLEPLEHRR 237  
Qy 175 RKIQKVVNTYTP-----GKKIWLEGVY--TTSAG-----GTNNLSDSY-----AAGF 214  
Db 238 RTICGTPNLYSPEVLNKGHCESDIWALGCVMTMLLGRPPFTTNLKYRCIREARY 297  
Qy 215 LWLNTL-----GMLANQ-----GIDVVRHSFFDHGY--NHLVDQNFNPLPDYWLS 258  
Db 298 TWPSSLLAPAKHLIASMLSKNPEDRPSLDDIIRHDFLQGTTPDRLSSCCCHTVPDFHLS 357  
Qy 259 LLYKRLIGPKVLAVHVGAGLQKPR----PGRVIRDKLRIYAHCTNNHHNNHNVVRSIT 311  
Db 358 SPARKNF-KKAAALFGGKDKKARYIDTHNRVSKEDDIY-----KLRLDLKKTST 408

RESULT 12  
US-09-505-744-2  
Sequence 2, Application US/09505744  
Patent No. 6245544  
GENERAL INFORMATION:  
APPLICANT: Karen M. Anderson  
APPLICANT: Mark M. Bouzyk  
APPLICANT: Michael J. Hansbury  
APPLICANT: Jeffrey R. Jackson  
APPLICANT: Sandhya S. Nerurkar  
APPLICANT: Amy K. Roshak  
TITLE OF INVENTION: HUMAN SERUM INDUCIBLE KINASE (SNK)  
FILE REFERENCE: GH-70231-D1  
CURRENT APPLICATION NUMBER: US/09/505,744  
CURRENT FILING DATE: 2000-02-16  
EARLIER APPLICATION NUMBER: 09/136,282  
EARLIER FILING DATE: 1998-08-20  
EARLIER APPLICATION NUMBER: 60/056,112  
EARLIER FILING DATE: 1997-08-20  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 685  
TYPE: PRT  
ORGANISM: HOMO SAPIENS  
US-09-505-744-2

Query Match 4.3%; Score 92.5; DB 4; Length 685;  
Best Local Similarity 23.2%; Pred. No. 0.43;  
Matches 83; Conservative 46; Mismatches 139; Indels 89; Gaps 18;

Qy 6 RGGPGPDYILKNYEDDIVRSVDALDKQKCKIAQHPDVMLELQREKAAQ-----M 55

Db 90 KGG-----FAKCYE---MTDLTNKVVAAKIIPIHSRVAKPHQREKIDKEIHLRHLHK 140  
Qy 56 HLVLKKEQFSNTYNNLIITPEPNNTYTHGRAVNGSOLGKDYIQLKSLQIP-IRIYSRASL 114  
Db 141 HVQFYHYFEDKENIYILLEYCSRSM-----AHILK---ARKVLTEPEVRYLQIV 190  
Qy 115 YGPNIGRPRKNVIALLDGFMKVGASTVDVATWQHCYIDGRVVKVMDFLKTRLLDLSQI 174  
Db 191 SGLKYLHEQE---ILHRDLKLGNNFFIN-----EAMELKVGDFGLAARLEPLEHRR 237  
Qy 175 RKIQKVVNTYTP-----GKKIWLEGVY--TTSAG-----GTNNLSDSY-----AAGF 214  
Db 238 RTICGTPNLYSPEVLNKGHCESDIWALGCVMTMLLGRPPFTTNLKYRCIREARY 297  
Qy 215 LWLNTL-----GMLANQ-----GIDVVRHSFFDHGY--NHLVDQNFNPLPDYWLS 258  
Db 298 TWPSSLLAPAKHLIASMLSKNPEDRPSLDDIIRHDFLQGTTPDRLSSCCCHTVPDFHLS 357  
Qy 259 LLYKRLIGPKVLAVHVGAGLQKPR----PGRVIRDKLRIYAHCTNNHHNNHNVVRSIT 311  
Db 358 SPARKNF-KKAAALFGGKDKKARYIDTHNRVSKEDDIY-----KLRLDLKKTST 408

RESULT 13  
US-08-435-675B-5  
Sequence 5, Application US/08435675B  
Patent No. 5710250  
GENERAL INFORMATION:  
APPLICANT: Ellis, Steven Bradley  
APPLICANT: Williams, Mark E.  
APPLICANT: Harpold, Michael Miller  
APPLICANT: Schwartz, Arnold  
APPLICANT: Brenner, Robert  
TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/435,675B  
FILING DATE: 05-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,083  
FILING DATE: 28-SEP-1994  
APPLICATION NUMBER: US 07/914,231  
FILING DATE: 13-JUL-1992  
APPLICATION NUMBER: US 07/603,751  
FILING DATE: 08-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-53193  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
TELEX:  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1106 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear



Search completed: July 30, 2002, 08:15:07  
Job time: 289 sec

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/223,305
; FILING DATE: April 4, 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/868,354
; FILING DATE: April 10, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/745,206
; FILING DATE: 13-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/620,250
; FILING DATE: 30-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/482,384
; FILING DATE: 20-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,751
; FILING DATE: 04-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US89/01408
; FILING DATE: 04-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/176,899
; FILING DATE: 04-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-52517
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)238-0999
; TELEFAX: (619)238-0062
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1103 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-455-543A-53

Query Match      4.08; Score 86; DB 1; Length 1103;
Best Local Similarity 18.7%; Pred. No. 4.7;
Matches 68; Conservative 54; Mismatches 120; Indels 122; Gaps 14;

Qy 124 KNVIALDGFMKVAGSTVDVAVTWQHCVYIDGRVVKVMDFLKTRLLDTLSD----- 172
Db 252 KDMILVD-----VSGSLKLIKRTSVSEMLETSLDDDFVNVASFNS 296

Qy 173 ---QIRKIQKVNTYTPGKIKWLEGVVTTSAGGTNNLSDSYAAGFLWL-----NT 219
Db 297 NAQDVSCFQHLVQANVRNKKVLDVANNITAKGITDYKKGFSFAFEQLLNNVSRANCNK 356

Qy 220 LGMLANOGIDVIRHSFSDHCYNHLY-----DQNFNPLPDYWLSL----- 259
Db 357 IIMLFTDGEERAQEIENKYNKKKKRVRFVSGQHNYERGIQWMACENKGYEIPSI 416

Qy 260 -----LYKRLIG-PRVLA-----VHVAGLQKRPGRVIRDKRIYAHCTN 299
Db 417 GAIRINTQEYLDVLGRPMVLGAKQVQWTVNVDLDALEL---GLVITGTLPVENITGQ 472

Qy 300 HHNNHYVRSITLPIINLHRSRKKIKLAGTLRDKLVHQLQYPQGBGLSKSVQLNQSP 359
Db 473 FENKTNLKNQLILGVMGVDVSLEDIK-----RLTPRFTLCPNGY----YFAIDPNGYV 521

Qy 360 LVMYDDGTLPPLKPRPLRAGTLPV-----PVTMGFFVVKVNNALA 401
Db 522 LLH-----PNLQPKPIGVG----IPTINLKRKRPNIQNPKSQEPVTLDFLDAELENDIK 571

Qy 402 CRYR 405
Db 572 VEIR 575
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:14:03 ; Search time 141.47 Seconds  
(without alignments)  
292.857 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_162\_534

Perfect score: 1956

Sequence: 1 KCCKTAQHPDVMLELQREKA.....PVTMGFFVKNVNALACRYR 373

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1956	100.0	534	22	AA85216 Heparanase-like pr
2	1949	99.6	492	22	AA84664 Amino acid sequenc
3	1917	98.0	592	22	AAU07424 Human heparanase-1
4	1917	98.0	592	22	AA85215 Heparanase-like pr
5	1917	98.0	592	22	AA85215 Heparanase-like pr
6	1907	97.5	592	22	AA85215 Heparanase-like pr
7	1743	89.1	439	22	AA851062 Human heparanase-2
8	1743	89.1	480	22	AAU07418 Human heparanase-1
9	1743	89.1	480	22	AAU07418 Novel human extrac
10	1743	89.1	480	22	AA85217 Heparanase-like pr
11	1740	89.0	538	22	AA85217 Human heparanase,

12	1112	56.9	214	22	AA85216
13	1112	56.9	214	22	AA85216
14	805	41.2	380	20	AA17085
15	791.5	40.5	543	20	AA02345
16	791.5	40.5	543	21	AA08849
17	791.5	40.5	543	21	AA57590
18	791.5	40.5	543	21	AA52990
19	791.5	40.5	543	22	AA57635
20	791.5	40.5	592	20	AA02346
21	791.5	40.5	592	21	AA08850
22	788.5	40.3	530	20	AA34173
23	788.5	40.3	543	22	AA88361
24	787.5	40.3	543	20	AA17082
25	787.5	40.3	543	22	AA86206
26	787.5	40.3	588	20	AA30124
27	784.5	40.1	535	21	AA08851
28	779	39.8	380	20	AA17084
29	778	39.8	156	22	AA65963
30	744.5	38.1	532	20	AA17083
31	485	24.8	262	22	AA24147
32	403	20.6	488	22	AA31469
33	388	19.8	488	22	AA31470
34	383	19.6	488	22	AA31472
35	375	19.2	488	22	AA31471
36	278	14.2	137	22	AA65961
37	278	14.2	159	22	AA65964
38	247	12.6	256	21	AA13479
39	213	10.9	112	22	AAU07425
40	203	10.4	118	22	AA65962
41	155	7.9	935	22	AB89219
42	144	7.4	38	20	AA34186
43	127	6.5	24	22	AA85220
44	127	6.5	25	22	AA85221
45	121	6.2	23	22	AA85219

#### ALIGNMENTS

RESULT 1

AA85216  
ID AA85216 standard; Protein; 534 AA.

AC AA85216;

DT 07-SEP-2001 (first entry)

DE Heparanase-like protein Hpa2 splice variant #2.

XX Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;  
KW cytosolic; neuroprotective; cerebrotectic; immunosuppressive;  
KW antipsoriatic; nootropic; antiinflammatory; antiarthritic; antiasthmatic;  
KW antidiabetic; antiarteriosclerotic; vulnary.

XX Homo sapiens.

XX WO200146392-A2.

XX 28-JUN-2001.

XX 21-DEC-2000; 2000WO-GB04963.

XX 22-DEC-1999; 99GB-0030392.

XX 07-APR-2000; 2000GB-0008713.

XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.

XX McKenzie EA, Stamps AC, Terrett JA, Tyson KL;

XX WPI; 2001-418056/44.

XX N-PSDB; AAH22672.

XX Novel homologs of heparanase, present in three splice variants, useful

PT for identifying agents that modulate heparanase, useful in the  
PT treatment and/or prophylaxis of abnormal levels of heparanase -  
XX  
PS  
XX  
XX Claim 1; Fig 2; 97pp; English.  
XX  
XX The invention provides a homologue to heparanase which is present in  
CC three splice variants. The heparanase homologue polypeptide is useful in  
CC the treatment of a human or non-human animal or for use in diagnosis.  
CC Vectors comprising the heparanase homologue polynucleotides are useful in  
CC the transformation or transfection of a prokaryotic or eukaryotic host.  
CC The modulators of the polypeptide are useful in the manufacture of a  
CC medicament for the treatment and/or prophylaxis of a condition/disease  
CC associated with abnormal levels of the heparanase homologue, including  
CC cancer, central nervous system (CNS) and neurodegenerative diseases,  
CC cardiovascular diseases such as stenosis following angioplasty and  
CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,  
CC allografts, inflammatory diseases, arthritis, vascular stenosis,  
CC tumour growth and progression, asthma, Alzheimer's disease, diabetic  
CC retinopathy, wound healing and inflammation. The polypeptide is also  
CC useful in diagnosis and research. The present sequence represents the  
CC amino acid sequence of the mid-sized splice variant of the heparanase-  
CC like protein Hpa2 of the invention.  
XX  
XX Sequence 534 AA;

Query Match 100.0%; Score 1956; DB 22; Length 534;  
Best Local Similarity 100.0%; Pred. No. 1.9e-199;  
Matches 373; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 KGCKTAQHPDVMLELQREKAAQMHLVLLKEQFSNTYSNLILTEPNRYTMHGRAVNGSQL 60  
Db 162 kyckiaqhpdvmlleqrekaqmhlvllkeqfsntysnliltepnrytmhgravgsql 221  
Qy 61 GKDYLQKSLQPIRIYRASLYGNIGRPNKRVIALDGFPMKVGSTVDVWQHCVYD 120  
Db 222 gkdylqkslqpriyrsaslygnigrpnknvialldgfmkvagstvdtwqhcyid 281  
Qy 121 GRVVKVMDFLKRLDLDLSDQIRKIQKVNTYTPGKKIWLGEVVTTSAGGTNNLSDSYAA 180  
Db 282 grvkvmdflkrlldtldsdqirkikqvntytpgkklwlegvvttsaggtnnlsdsyaa 341  
Qy 181 GFLWLTLMLANOGIDVYIRHSFDDHGYNHLVDQNFNPLDPYWLISLKYKRLIGPKVLAV 240  
Db 342 gflwltlmglangidvyrhsfddhgynhlvdqnfpldpdywlsllykrligpkvlav 401  
Qy 241 HVAGLQRPGRVIRDKLRIYAHCTNNHNNHYVRSSTLFTIINLHRSRKKIKLAGTLRD 300  
Db 402 hvaglrqprgrvirdklriyahctnnhnnhnyvrgstlftiinlhrsrrkkiklagtlrd 461  
Qy 301 KLVHGYLLQPYGQGLKSKSVOLNQPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFF 360  
Db 462 klvhgyllqpygqglksksvqlngqplvmvddgtlpeklprrlagrtlviptvtmgff 521  
Qy 361 VVKNNALACRYR 373  
Db 522 vvknnalacryr 534

RESULT 2  
AAB84664  
ID AAB84664 standard; Protein; 492 AA.  
XX AAB84664;  
AC AAB84664;  
XX  
DT 05-SEP-2001 (first entry)  
DE Amino acid sequence of human heparanase-like polypeptide.  
XX  
KW Human; heparanase-like polypeptide; gene therapy; cancer; angiogenesis;  
KW trauma; autoimmune disease; skin disease; cardiovascular disease;  
KW nervous system disease; inflammation; arthritis; genitalia;  
KW male fertility; erectile dysfunction.

XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FT Misc-difference 407 /note= "unspecified residue encoded by KCA"  
PN  
XX WO200148161-A2.  
XX  
XX 05-JUL-2001.  
XX  
XX 18-DEC-2000; 2000WO-EP12909.  
XX  
XX 23-DEC-1999; 99EP-0125831.  
XX (SCHD ) SCHERING AG.  
XX  
XX Siemeister G, Weiss B;  
XX  
XX WPI: 2001-418259/44.  
DR N-PSDB; AAH28347.  
XX  
XX Human Heparanase-like polynucleotide encoding polypeptides useful for  
PT modulating expression of the polypeptide and for treating cancer,  
PT cancer metastasis, aberrant angiogenesis by gene therapy technique -  
XX  
XX Claim 9; Page 30; 30pp; English.  
XX  
XX The present sequence represents a human heparanase-like polypeptide.  
CC Heparanase-like polynucleotides are useful as a source of probes,  
CC primers and antisense molecules, and in gene therapy. Heparanase-like  
CC polynucleotides and polypeptides are useful for treating several  
CC disorders e.g., cancer, cancer metastasis. The oligonucleotides are  
CC also useful as diagnostic markers for the diagnosis of disorder such  
CC as cancer, cancer metastasis and aberrant angiogenesis. They may also  
CC act as diagnostic markers for diagnosis of disorder such as cancer,  
CC cancer metastasis and aberrant angiogenesis. The heparanase polypeptides  
CC and polynucleotides are also useful for treating trauma, autoimmune  
CC diseases, skin diseases, cardiovascular diseases, nervous system  
CC diseases, and inflammation including arthritis. Since the polynucleotide  
CC is preferentially expressed in male genitalia, modulation of its  
CC expression and/or activity may be used for medical intervention in male  
CC genitalia function that is male fertility control, erectile dysfunction.  
XX  
XX Sequence 492 AA;

Query Match 99.6%; Score 1949; DB 22; Length 492;  
Best Local Similarity 99.5%; Pred. No. 9.6e-199;  
Matches 371; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 KGCKTAQHPDVMLELQREKAAQMHLVLLKEQFSNTYSNLILTEPNRYTMHGRAVNGSQL 60  
Db 120 kyckiaqhpdvmlleqrekaqmhlvllkeqfsntysnliltepnrytmhgravgsql 179  
Qy 61 GKDYLQKSLQPIRIYRASLYGNIGRPNKRVIALDGFPMKVGSTVDVWQHCVYD 120  
Db 180 gkdylqkslqpriyrsaslygnigrpnknvialldgfmkvagstvdtwqhcyid 239  
Qy 121 GRVVKVMDFLKRLDLDLSDQIRKIQKVNTYTPGKKIWLGEVVTTSAGGTNNLSDSYAA 180  
Db 240 grvkvmdflkrlldtldsdqirkikqvntytpgkklwlegvvttsaggtnnlsdsyaa 299  
Qy 181 GFLWLTLMLANOGIDVYIRHSFDDHGYNHLVDQNFNPLDPYWLISLKYKRLIGPKVLAV 240  
Db 300 gflwltlmglangidvyrhsfddhgynhlvdqnfpldpdywlsllykrligpkvlav 359  
Qy 241 HVAGLQRPGRVIRDKLRIYAHCTNNHNNHYVRSSTLFTIINLHRSRKKIKLAGTLRD 300  
Db 360 hvaglrqprgrvirdklriyahctnnhnnhnyvrgstlftiinlhrsrrkkiklagtlrd 419  
Qy 301 KLVHGYLLQPYGQGLKSKSVOLNQPLVMVDDGTLPELKPRLRAGRTLVIPTVTMGFF 360  
XX  
XX



Db 420 klvhqyllqpyggqgkkskvsqngqplvmvddgtlpeklprprlagrtlvipptmgfy 479

Qy 361 VVKVNALACRYR 373

Db 480 vkvvnalacryr 492

RESULT 3

AAU07424

XX AC AAU07424;

XX DE 18-DEC-2001 (first entry)

XX DE Human heparanase-like protein splice variant #1.

XX KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;

XX KW antiproliferative; cardiac; vasotropic; cerebroprotective; neurotropic;

XX KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;

XX KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;

XX KW hyperproliferative disorder; neoplasm; cardiovascular disorder;

XX KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;

XX KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;

XX KW wound healing; food additive; heparanase.

XX OS Homo sapiens.

XX PN WO200179253-A1.

XX PN 25-OCT-2001.

XX XX 11-APR-2001; 2001WO-US11643.

XX XX 18-APR-2000; 2000US-198123P.

XX XX (HUMA-) HUMAN GENOME SCI INC.

XX XX Fiscella M, Shi Y, Ebner R, Ruben SM;

XX XX WPI; 2001-611720/70.

XX XX N-PSDB; AAS13848.

XX XX New nucleic acids encoding extracellular matrix polypeptides, for

XX XX diagnosing, treating, preventing or ameliorating human disorders and

XX XX disease, such as, autoimmune, hyperproliferative or cardiovascular

XX XX disorders -

XX XX Disclosure; Page 14; 308pp; English.

XX CC The invention relates to novel isolated polynucleotides (I) encoding

XX CC extracellular matrix (ECM) polypeptides. (I) and a polypeptide encoded by

XX CC (I) are used to prevent, treat or ameliorate a medical condition in e.g.

XX CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They

XX CC are also used in diagnosing a pathological condition or susceptibility to

XX CC a pathological condition. The antibodies to the polypeptides can also be

XX CC used in alleviating symptoms associated with the disorders and in

XX CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked

XX CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated

XX CC include autoimmune diseases e.g. rheumatoid arthritis,

XX CC hyperproliferative disorders e.g. neoplasms of the breast or liver,

XX CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders

XX CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.

XX CC Alzheimer's disease, infections caused by bacteria, viruses and fungi

XX CC and ocular disorders e.g. corneal infection. The polypeptides can also

XX CC be used to aid wound healing and epithelial cell proliferation, to

XX CC prevent skin aging due to sunburn, to maintain organs before

XX CC transplantation, for supporting cell culture of primary tissues, to

XX CC regenerate tissues and in chemotaxis. The polypeptides can also be used

XX CC as a food additive or preservative to increase or decrease storage

XX CC capabilities. The present sequence represents the amino acid sequence

XX CC of human heparanase-like protein, splice variant #1.

SQ Sequence 592 AA;

Query Match 98.0%; Score 1917; DB 22; Length 592;

Best Local Similarity 86.5%; Pred. No. 3.3e-195;

Matches 373; Conservative 0; Mismatches 0; Indels 58; Gaps 1;

Qy 1 KGCKIAQHPDVNMLEQREKAAQMHLVLKQFSTNTYSNLILT----- 42

Db 162 kgckiaqhpdvmlqlrekaaqmhlvllkqfntysnliltarsldklynfadcsglhl 221

Qy 43 -----EPNNYRTMHGRAVNGSGLGK 62

Db 222 ifalnairpnpswnsssalilkyssaskyniswelgnepnnyrtmhgravngsqlgk 281

Qy 63 DYIOLKSELQPIRYVSRLSYGNIGRPRKNVIALLDGFMKVAGSTVDVAVTWOHCYIDGR 122

Db 282 dyiolksllqpirlyrsaslygpnigrprknvialldgfmkvagstvdavtwqhcylodr 341

Qy 123 VVKVMDFLKTRLLDLSQIRKIQVNTYTPGKKIWIIEGVVTTTSAGGTNNLSDSYAAGF 182

Db 342 vkvvmdflktrlldlsdqirkiqkvntypgkkiwilegvvtttsaggtnnlsdsyaagf 401

Qy 183 LWLNTGLMANOGIDVIRHSFFDHGYNHLVDQNFNPDPYWLSSLYKRLIGPKVLAVHV 242

Db 402 lwlntlgmlanogidvvrhsffdhgynhlvdqnfnpdpywlslykrligpkvlavhv 461

Qy 243 AGLORKPGRGVIRDKLRIYAHCTNHHNHNVRGSITLFIINLHRSKKIKLAGTLRDKL 302

Db 462 aglorkpgrgvirdklriyahctnhhnhnvrgsitlfiinlhrskkiklagtlrddl 521

Qy 303 VHQYLLQPYGQGLKSKSVQLNGOPLVMVDGTLPELKPRLRAGRTLVIPTVMGFVW 362

Db 522 vhyllpygqeglkksvqlngqplvmvddgtlpeklprprlagrtlviptvmgfvw 581

Qy 363 KVNVALACRYR 373

Db 582 knvnalacryr 592

RESULT 4

AAB85215

ID AAB85215 standard; Protein; 592 AA.

XX AC AAB85215;

XX DT 07-SEP-2001 (first entry)

XX DE Heparanase-like protein Hpa2 splice variant #1.

XX KW Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;

XX KW cytosolic; neuroprotective; cerebroprotective; immunosuppressive;

XX KW antipsoriatic; nootropic; antiinflammatory; antiarthritic; antiasthmatic;

XX KW antidiabetic; antiarteriosclerotic; vulnerary.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Misc-difference 237

XX FT /label= unknown

XX FT /note= "encoded by ANC"

XX PN WO200146392-A2.

XX PD 28-JUN-2001.

XX XX 21-DEC-2000; 2000WO-GB04963.

XX XX 22-DEC-1999; 99GB-0030392.

XX XX 07-APR-2000; 2000GB-0008713.

XX XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.



Qy 303 VHQYLLQPYGQGLKSKSVQVNGQPLVMVDGTLDELKPRPLRAGRTLVIPTVTMGFFVW 362  
Db 522 VHQYLLQPYGQGLKSKSVQVNGQPLVMVDGTLDELKPRPLRAGRTLVIPTVTMGFFVW 581  
Qy 363 KVNALACRYR 373  
Db 582 knvnalacryr 592

RESULT 6  
AAB81062  
ID AAB81062 standard; Protein; 592 AA.  
XX AAB81062;  
AC AAB81062;  
XX 20-JUN-2001 (first entry)  
DT 20-JUN-2001 (first entry)  
DE Human Heparanase-2 amino acid sequence.  
XX Heparanase 2; human; endoglucuronidase; heparan sulphate; metastasis;  
KW neangiogenesis; vaccine; autoimmune disorder; blood coagulation;  
KW cancer; diabetes; ischaemia; sepsis; stroke; cardiovascular; thrombosis.  
XX OS Homo sapiens.  
XX Key Location/Qualifiers  
FT Region 156..169  
FT /label= Immunogenic\_epitope  
FT Region 249..262  
FT /label= Immunogenic\_epitope  
FT Region 505..518  
FT /label= Immunogenic\_epitope  
XX WO200121814-A1.  
XX 29-MAR-2001.  
XX 11-SEP-2000; 2000WO-EP08837.  
XX 23-SEP-1999; 99EP-0118805.  
XX 07-JUL-2000; 2000EP-0114649.  
XX (MERE ) MERCK PATENT GMBH.  
XX Duecker K, Sirrenberg C;  
XX WPI; 2001-308089/32.  
XX N-PSDB; AAF86101.  
XX New heparanase-2 polypeptide useful in diagnosing (the susceptibility  
of a subject to) and as vaccines against e.g. autoimmune disorders,  
cardiovascular disease, cancer, diabetes, ischemia, sepsis, stroke, or  
thrombosis -  
XX Claim 1; Page 42-43; 46pp; English.  
XX This invention relates to a human heparanase-2 protein and the cDNA  
sequence encoding it. Heparanase-2 is a member of the endoglucuronidase  
family of polypeptides and it degrades heparan sulphate proteoglycans  
HSPGs (ubiquitous macromolecules of cell surfaces, basement membranes and  
the extracellular matrix). HSPGs support the vascular endothelium and  
stabilise the structure of the capillary wall. Heparanases may be  
associated with neangiogenesis and metastasis related to malignant  
tumour formation. Heparanase-2 polynucleotides and proteins are useful as  
vaccines for inducing an immunological response against autoimmune  
disorders, blood coagulation disorders, cancer, diabetes, ischaemia,  
sepsis, stroke, cardiovascular diseases, or thrombosis, as well as in  
diagnosing (the susceptibility of a subject to) these diseases.  
Heparanase-2 fragments may be used as immunogens to produce antibodies  
immunospesific to the polypeptides, and to identify membrane bound  
soluble receptors, agonists or antagonists that compete with the binding  
of the polypeptide to the receptors. An antibody specific for  
heparanase-2 can be used in the diagnosis of the above diseases and in

CC isolating or identifying clones expressing heparanase-2. The present  
CC sequence represents heparanase-2. Three regions of heparanase-2 with high  
CC immunogenicity (immunogenic epitopes) can be used to raise antibodies  
CC against heparanase-2.  
XX Sequence 592 AA;  
Qy 1 KGCKIAQHPVMLEQREKAAQMHVLVLLKEQFSNTYSNLILT----- 42  
Db 162 kgckiaqhpdvmlvlgrekaqmhvlvllkeqfsntysnliltcarsldklynfadcsghl 221  
Qy 43 -----EPNNYRTMHGRAVNGSOLGK 62  
Db 222 ifalnalmrnpunssalsllkysaskkynlswelgneppnyrtmhgravnsglqk 281  
Qy 63 DYIQLKSLIQPIRIYSRASLYGNIGRPKNVIALLDGFMKVAGSTVDVAVTWHQCYIDGR 122  
Db 282 dyiqklsllqpiriysraslygnigrknviaalldgfmkvagstvdaavtwhqcyidgr 341  
Qy 123 VVKVMDFLKRLDLSQIRKIQKVNTYTPGKKIWLEGVVTTTSGGTNNISDSYAAGF 182  
Db 342 vvkvmdfklrldtldsqirkiqkvntytpgkkiwlegvvttsggtnnlsdsyaagf 401  
Qy 183 LWLNTLGLMANGODVIRHSEFDHGYNHLVDONENPLPDYWLSSLYKRLIGPKVLAVHV 242  
Db 402 lwintlglmangodvirkhsfddhgynhlvdqnlfpdywlslylkrllgpkvlavhv 461  
Qy 243 AGLQRKPRGRVIRDKLRIYAHCTNHHNHYVGRSTTFLIINLHRSRKKIKLAGTLRDLK 302  
Db 462 aglqrkprgrvirdklriyahcthhnhmyvrgsitifilnlhrrskkiklagtlrldkl 521  
Qy 303 VHQYLLQPYGQGLKSKSVQVNGQPLVMVDGTLDELKPRPLRAGRTLVIPTVTMGFFVW 362  
Db 522 VHQYLLQPYGQGLKSKSVQVNGQPLVMVDGTLDELKPRPLRAGRTLVIPTVTMGFFVW 581  
Qy 363 KVNALACRYR 373  
Db 582 knvnalacryr 592

RESULT 7  
AAU07423  
ID AAU07423 standard; Protein; 439 AA.  
XX AAU07423;  
AC AAU07423;  
XX 18-DEC-2001 (first entry)  
DE Human heparanase-like protein.  
XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;  
KW antiproliferative; cardiant; vasotropic; cerebroprotective; nootropic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW extracellular matrix; ECM; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;  
KW nervous system disorder; Alzheimer's disease; ocular disorder; sunburn;  
wound healing; food additive; heparanase.  
XX Homo sapiens.  
XX WO200179253-A1.  
XX 25-OCT-2001.  
XX 11-APR-2001; 2001WO-US11643.  
XX 18-APR-2000; 2000US-198123P.



QY 103 KVGSTVDVAVTWHQCYIDGRVVKVMDFLKTRLLDLSQIRKIQKVVNTYTPGKKIWLGG 162  
Db 210 kvagstvdavtwqhcycidgrvkvkvmfklrlltdlsdqirkiqkvntypgkklwleg 269  
QY 163 VVTSAGTNNLSDSYAAGFLWNLTLGMLANOGIDVIRHSHFFDGHYNHLVDONENPLPD 222  
Db 270 vvttsaggtnnlsdsyaagflwntlgmlanogidvirhsffdhgynhlvdqnfnpdp 329  
QY 223 YWLSLLYKRLIGPKVLAVHVAGLQRPGRVIRDKLRIYAHCTNNHNNHYRGSITLFI 282  
Db 330 ywlslllykrligpkvlavhvaglrkprgvrirdklriyahctnnhnnhyrgsitlfi 389  
QY 283 INLHRSRKKIKLAGTLRDKLVHVOYLLOPYGQGLSKSVQLNGQPLVWVDGDTLPKLPKR 342  
Db 390 inlhrrskkiklagtlrdklvhqyllyqpygqglksksvqlngqplvmvddgtlpelkpr 449  
QY 343 PLRAGRTLVIPIPPVTMGFFVVKNNVALACRYR 373  
Db 450 plragrtlviippvtmgffvkvnnvalacryr 480

RESULT 9  
AAB85217  
ID AAB85217 standard; Protein; 480 AA.  
AC AAB85217;  
XX  
DT 07-SEP-2001 (first entry)  
XX  
DE Heparanase-like protein Hpa2 splice variant #3.  
XX  
KW Heparanase; splice variant; homologue; heparanase-like protein; Hpa2;  
KW cytosolic; neuroprotective; cerebroprotective; immunosuppressive;  
KW antiproliferative; nontoxic; antiinflammatory; antiarthritic; antiasthmatic;  
KW antidiabetic; antiarteriosclerotic; vulnerary.  
XX  
OS Homo sapiens.  
XX  
PN WO200146392-A2.  
XX  
PD 28-JUN-2001.  
XX  
PF 21-DEC-2000; 2000WO-GB04963.  
XX  
PR 22-DEC-1999; 99GB-0030392.  
PR 07-APR-2000; 2000GB-0008713.  
XX  
PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
XX  
PI McKenzie EA, Stamps AC, Terrett JA, Tyson KL;  
XX WPI; 2001-418056/44.  
DR N-PSDB; AAB22673.  
XX  
PT Novel homologs of heparanase, present in three splice variants, useful  
PT for identifying agents that modulate heparanase, useful in the  
PT treatment and/or prophylaxis of abnormal levels of heparanase  
XX  
PS Claim 1; Fig 3; 97pp; English.  
XX

CC The invention provides a homologue to heparanase which is present in  
CC three splice variants. The heparanase homologue polypeptide is useful in  
CC the treatment of a human or non-human animal or for use in diagnosis.  
CC Vectors comprising the heparanase homologue polynucleotides are useful in  
CC the transformation or transfection of a prokaryotic or eukaryotic host.  
CC The modulators of the polypeptide are useful in the manufacture of a  
CC medicament for the treatment and/or prophylaxis of a condition/disease  
CC associated with abnormal levels of the heparanase homologue, including  
CC cancer, central nervous system (CNS) and neurodegenerative diseases,  
CC cardiovascular diseases such as stenosis following angioplasty and  
CC atherosclerosis, autoimmune diseases, psoriasis, lupus erythematosus,  
CC allografts, inflammatory diseases, arthritis, vascular stenosis,  
CC tumour growth and progression, asthma, Alzheimer's disease, diabetic

CC retinopathy, wound healing and inflammation. The polypeptide is also  
CC useful in diagnosis and research. The present sequence represents the  
CC amino acid sequence of the smallest splice variant of the heparanase-  
CC like protein Hpa2 of the invention.  
XX  
SQ Sequence 480 AA;  
Query Match 89.1%; Score 1743; DB 22; Length 480;  
Best Local Similarity 100.0%; Pred. No. 8.7e-177;  
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 43 EPNVNTMHGRAVNSQLGKDYIQLKSLQPIRIYSRASLYGNIGRPRKNNVIALDGM 102  
Db 150 epnvntmhgravnsqigkdyiqkllqpiriysraslygnigrprknnvialldgm 209  
QY 103 KVGSTVDVAVTWHQCYIDGRVVKVMDFLKTRLLDLSQIRKIQKVVNTYTPGKKIWLGG 162  
Db 210 kvagstvdavtwqhcycidgrvkvkvmfklrlltdlsdqirkiqkvntypgkklwleg 269  
QY 163 VVTSAGTNNLSDSYAAGFLWNLTLGMLANOGIDVIRHSHFFDGHYNHLVDONENPLPD 222  
Db 270 vvttsaggtnnlsdsyaagflwntlgmlanogidvirhsffdhgynhlvdqnfnpdp 329  
QY 223 YWLSLLYKRLIGPKVLAVHVAGLQRPGRVIRDKLRIYAHCTNNHNNHYRGSITLFI 282  
Db 330 ywlslllykrligpkvlavhvaglrkprgvrirdklriyahctnnhnnhyrgsitlfi 389  
QY 283 INLHRSRKKIKLAGTLRDKLVHVOYLLOPYGQGLSKSVQLNGQPLVWVDGDTLPKLPKR 342  
Db 390 inlhrrskkiklagtlrdklvhqyllyqpygqglksksvqlngqplvmvddgtlpelkpr 449  
QY 343 PLRAGRTLVIPIPPVTMGFFVVKNNVALACRYR 373  
Db 450 plragrtlviippvtmgffvkvnnvalacryr 480

RESULT 10  
AAY97634  
ID AAY97634 standard; Protein; 480 AA.  
XX  
AC AAY97634;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE Human heparanase, hnhp1 pn5 form, protein sequence.  
XX  
KW Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;  
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;  
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;  
KW gene therapy; human.  
XX  
OS Homo sapiens.  
XX  
PN WO200100643-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 19-JUN-2000; 2000WO-IL00358.  
XX  
PR 25-JUN-1999; 99US-0140801.  
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
XX  
PI Pecker I, Michal I, Itzhaki H;  
XX WPI; 2001-137930/14.  
DR N-PSDB; AAA91099.  
XX  
PT New polynucleotides and polypeptides that are distantly homologous to  
PT heparanase, useful in wound healing, as well as in gene therapy  
PT protocols for angiogenesis, restenosis, atherosclerosis, or  
PT inflammation -





DR WPI; 2001-465569/50.  
DR N-PSDB; AA198878.  
XX Isolated nucleic acid molecule encoding excretory system antigen is  
PT used in preventing, treating or ameliorating a medical condition -  
PT  
XX  
PS Claim 11; SEQ ID NO 642; 574pp + Sequence Listing; English.  
XX  
CC The invention relates to novel excretory system related human  
CC polynucleotides (AA198567-AA199503) and the encoded proteins  
CC (AA199594-AA199913) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy, especially  
CC disorders related to the excretory system. The genes are isolated  
CC from a range of human tissues disclosed in the specification. The  
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 214 AA;

Query Match 56.9%; Score 1112; DB 22; Length 214;  
Best Local Similarity 100.0%; Pred. No. 5.1e-110; Indels 0; Gaps 0;  
Matches 210; Conservative 0; Mismatches 0;

QY 164 VTTSGAGTNNLSDSYAAGFLWLTGLMGLANQGDVIRHSFFDHGYNHLVDQNFNPLPDY 223  
Db 5 vttsaggtnnlstdsyagflwltgmlanqgdvirsffdhgynhlvdqnfnpdpdy 64  
QY 224 WLSLLYKRLIGPKVLAVHVGLOKRPGRVIRDKLRIYAHCTNHHNHYVGRSITLFI 283  
Db 65 wlsllkrligpkvlavhvaglokrpgrvirdklriyahctnhhnhnyvgrsitlfi 124  
QY 284 NLHRSRKKIKLAGTLRDKLVHQLVLLQPYGOEGLKSKSVOLNGOPLVMVDGTLPELKPRP 343  
Db 125 nlhrsrrkkiklagtlrdklvhlvllqpygoeglkksvqlngoplvmdvgtlpelkprp 184  
QY 344 LRAGRTLVIPTVTMGFFVKNVKNALACRYR 373  
Db 185 lragrtlviptvtmgffvknvnalacryr 214

RESULT 13  
AA198878  
ID AA198878 standard; Protein; 214 AA.  
XX  
AC AA198878;  
XX  
XX 24-OCT-2001 (first entry)  
DE Human bladder antigen, SEQ ID NO: 98.  
XX  
XX Human; bladder antigen; cytostatic; immunosuppressive; nootropic;  
KW neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;  
KW antiinflammatory; antitumor; vulnerary; anticonvulsant; antibacterial;  
KW antifungal; antiparasitic; cardiant; gene therapy; cancer;  
KW immune disorder; cardiovascular disorder; wound healing; infection;  
KW neurological disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200159064-A2.

XX  
PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01342.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214986.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216847.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0232514.  
PR 14-AUG-2000; 2000US-0232566.  
PR 14-AUG-2000; 2000US-0232567.  
PR 14-AUG-2000; 2000US-0232568.  
PR 14-AUG-2000; 2000US-0232570.  
PR 14-AUG-2000; 2000US-0232547.  
PR 14-AUG-2000; 2000US-0232575.  
PR 14-AUG-2000; 2000US-0232578.  
PR 14-AUG-2000; 2000US-0232579.  
PR 18-AUG-2000; 2000US-0232629.  
PR 22-AUG-2000; 2000US-0232681.  
PR 22-AUG-2000; 2000US-0232686.  
PR 22-AUG-2000; 2000US-0232718.  
PR 23-AUG-2000; 2000US-0232709.  
PR 30-AUG-2000; 2000US-02328924.  
PR 01-SEP-2000; 2000US-0232987.  
PR 01-SEP-2000; 2000US-0232987.  
PR 01-SEP-2000; 2000US-0233443.  
PR 01-SEP-2000; 2000US-0233444.  
PR 01-SEP-2000; 2000US-0233445.  
PR 05-SEP-2000; 2000US-0233459.  
PR 05-SEP-2000; 2000US-0233513.  
PR 06-SEP-2000; 2000US-0233437.  
PR 06-SEP-2000; 2000US-0233438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0234984.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.



PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2001US-0259678.  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX

DR WPI; 2001-514652/56.  
DR N-PSDB; AAI64065.  
XX  
PT Forty five bladder related polynucleotides, useful in the prevention,  
PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
PT disorders and neurological diseases -  
XX  
PS Claim 11; SEQ ID NO 98; 482pp + sequence listing; English.  
XX  
CC The invention relates to forty five novel bladder related  
CC polynucleotides. The polynucleotides and the polypeptides that they  
CC encode are useful in the diagnosis, treatment and prevention of:  
CC cancer, particularly breast and ovarian cancer, and other cancers  
CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal  
CC tract, liver, lung, or urogenital system; immune disorders such as  
CC Addison's disease, allergies, autoimmune haemolytic anaemia,  
CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple  
CC sclerosis, rheumatoid arthritis and ulcerative colitis; cardiovascular  
CC disorders such as myocardial ischaemias; wound healing; neurological  
CC diseases such as cerebral anoxia and epilepsy; and infectious diseases  
CC such as viral, bacterial, fungal and parasitic infections. Numerous  
CC examples of each type of disorder are given in the specification.  
CC The polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities. The polynucleotides are  
CC useful for chromosome identification. They are also useful as probes  
CC for diagnosing or treating a disorder related to the female reproductive  
CC system, particularly breast and/or ovary cancer. The present  
CC sequence is a novel bladder antigen provided in the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 214 AA:  
  
Query Match 56.9%; Score 1112; DB 22; Length 214;  
Best Local Similarity 100.0%; Pred. No. 5.le-110;  
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 164 VTTSGGTTNNLSDSYAAGFLWLTGLMLANQGDVIRHSFFDHGYNHLVDQNFNPLPDY 223  
|||||  
Db 5 vttsaggtnnlstdsyaagflwltglmlanqgdvirsffdhgynhlvdqnfplpdy 64  
|||||  
QY 224 WLSLYKRLIGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNHHNNHNVYRGSITLFI 283  
|||||  
Db 65 wisilykrligpkvlavhvaglqkprgrvirdkiriayahctnhhnhnyvrgsitfii 124  
|||||  
QY 284 NLHRSRKKIKLAGTLRDKLVHQYLLQPYGEGLSKSVQLNGOPLVMVDGDTLPKPRP 343  
|||||  
Db 125 nlhrsrrkkiklagtlrdklvhqyllqpyggeglsksvqlngoplvmdvddgtlpeikprp 184  
|||||  
QY 344 LRAGRTLVIPTVTMGFFVKNVKNALACRYR 373  
|||||  
Db 185 lragrtlivpptvmgtmgffvknvnalacryr 214  
  
RESULT 14  
AAAY17085  
ID AAY17085 standard; Protein; 380 AA.  
XX  
AC AAY17085;  
XX  
DT 21-JUL-1999 (first entry)  
XX  
DE Rat heparanase enzyme.  
XX  
KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;  
KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;  
KW arteriosclerosis; atherosclerosis; inflammation; tissue development;  
KW rat; HSPG.  
XX  
OS Rattus sp.  
XX



Qy 256 RDKLRIVAHCTNHHNHYVRGSITLFIINLHRSRKKIKLAGTLRDKLVHQYLLQPYQEG 315  
Db 430 --kirvylhctntdnprykegdltiyainlhavtkyrlrlypfsnkqvdkyllrplgphg 487  
Qy 316 LKSKSVOLNGOPLVMVDDGTLPCLKPRPLRAGRTLVIIPPVTMGFEVVKVNNALAC 370  
Db 488 llksvqlngltlkmvddqtlpplmekplrp9sslglpafsysffivrnakvaac 542

Search completed: July 30, 2002, 08:14:04  
Job time: 341 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:16:32 ; Search time 71.68 Seconds  
(without alignments)  
500.018 Million cell updates/sec

Title: us-09-836-461-2\_copy\_162\_534  
Perfect score: 1956  
Sequence: 1 KCKTKIAQHPDVMLELQREKA.....PVTMGFFVKNVNALACRYR 373

Scoring table: BLOSOM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	1743	89.1	480 2 JC7506	heparanase protein
2	301	15.4	521 2 T45608	hypothetical prote
3	155	7.9	190 2 T01953	hypothetical prote
4	102.5	5.2	1260 2 T14022	reverse transcript
5	101	5.2	535 2 S58740	cytochrome-c oxida
6	100.5	5.1	630 2 C71374	probable glucose i
7	95	4.9	512 2 T11261	cytochrome-c oxida
8	94.5	4.8	460 2 G90554	p46-like (mycoplas
9	94	4.8	625 2 G86855	glucose inhibited
10	94	4.8	1517 2 B81393	DNA-directed RNA p
11	93	4.8	320 2 D90478	conserved hypothet
12	92.5	4.7	328 2 AC1210	drdp-D-glucose 4,6
13	92.5	4.7	876 2 G89952	DNA polymerase I
14	92.5	4.7	2241 2 S09811	hypothetical prote
15	92	4.7	461 1 S65187	GPI-anchor biosynt
16	92	4.7	534 2 S17993	cytochrome-c oxida
17	91.5	4.7	335 2 E84992	tryptophan--trna l
18	90	4.6	650 2 F91654	regulatory protein
19	90	4.6	709 2 T16584	hypothetical prote
20	90	4.6	1616 2 T16600	vitellogenin vit-1
21	89.5	4.6	898 2 B8471	cytoplasmic aconit
22	88.5	4.5	286 2 A86576	chromosome partiti
23	88.5	4.5	286 2 B72047	chromosome partiti
24	88.5	4.5	341 2 D70761	probable fatty-acy
25	88.5	4.5	456 2 T24442	hypothetical prote
26	88.5	4.5	1272 2 C90593	hypothetical prote
27	88	4.5	555 2 AE2147	cytochrome c oxida
28	88	4.5	1106 1 CHRBA2	calcium channel pr
29	87.5	4.5	527 2 S44081	l1s globulin - oat

ALIGNMENTS

RESULT 1

JC7506  
heparanase protein 2a - human  
C:Species: Homo sapiens (man)  
C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 01-Dec-2000  
C:Accession: JC7506  
R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hirccock, M.;  
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000  
A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase fami  
A:Reference number: JC7506  
A:Accession: JC7506  
A:Molecule type: mRNA  
A:Residues: 1-480 <MCK>  
A:Cross-references: GB:AF282885  
C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and th  
erapies.  
C:Genetics:  
A:Gene: hpa2a  
A:Map position: 10q23-10q24  
C:Keywords: heparin binding; membrane bound

Query Match		89.1%	Score 1743;	DB 2;	Length 480;
Best Local Similarity		100.0%	Pred. No. 4.8e-138;		
Matches 331;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	43	EPNNRTMHGRAVNSQLGKDYIQLKSLLOPIRIYSRASLYGPNIGRPRKNVIALLDGFM	102		
Db	150	EPNNRTMHGRAVNSQLGKDYIQLKSLLOPIRIYSRASLYGPNIGRPRKNVIALLDGFM	209		
Qy	103	KVAGSTVDATVQHCYIDGRVVKVMDFLKTRLLDLSDDIRKIQKVVNTYTPGKKIWLGE	162		
Db	210	KVAGSTVDATVQHCYIDGRVVKVMDFLKTRLLDLSDDIRKIQKVVNTYTPGKKIWLGE	269		
Qy	163	VVTSAGGTNNLSDSYAAGFLWLNTLGLMLANQIDVIRHSFFDHGYNHLVDQNFPLPD	222		
Db	270	VVTSAGGTNNLSDSYAAGFLWLNTLGLMLANQIDVIRHSFFDHGYNHLVDQNFPLPD	329		
Qy	223	YWLSSLYKRLIGPKVLAVHAGVAGLQRPGRVIRKRIYAHCTNHHNNHYRGSTITLFI	282		
Db	330	YWLSSLYKRLIGPKVLAVHAGVAGLQRPGRVIRKRIYAHCTNHHNNHYRGSTITLFI	389		
Qy	283	INLHRSRKIKIAGTLRDLKLVHQLLOPYQBELSKSVOLANGQPLVMDVDDGTLPKLKPR	342		
Db	390	INLHRSRKIKIAGTLRDLKLVHQLLOPYQBELSKSVOLANGQPLVMDVDDGTLPKLKPR	449		
Qy	343	PLRAGRTLVIPPTMTGFFVYVKNVNALACRYR	373		
Db	450	PLRAGRTLVIPPTMTGFFVYVKNVNALACRYR	480		

RESULT 2

## T45608

hypothetical protein F13G24.30 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T45608  
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba  
submitted to the Protein Sequence Database, December 1999  
A:Reference number: Z23009  
A:Accession: T45608  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-521 <BEV>  
A:Cross-references: EMBL:AL133421  
A:Experimental source: cultivar Columbia; BAC clone F13G24  
C:Genetics:  
A:Map position: 5  
A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3  
A:Note: F13G24.30

Query Match 15.4%; Score 301; DB 2; Length 521;  
Best Local Similarity 28.2%; Pred. No. 3.4e-17;  
Matches 105; Conservative 50; Mismatches 127; Indels 90; Gaps 15;

Qy	52	GRAVNGSOLKDYIQLKSLLOPRTIRYVSRAISLYCPNIGRPKNVIALLDGF-----M	102
Dd	186	GASVSAELYGKDLTLVLKDVIN--KVYKNSWLHKRPIILAVP-----GGFYEQQWYTLL	235
Qy	103	KVAG-SIVDAVTWHQCXIDGR-----VKVMDFLKRLLDITLSDIRIKQKVVNTYTPG	155
Dd	236	EISGPSVDVVYTHHYNLGSGNDPALVKIMD---PSYLSQVSKTFKDYNQTIOEHGP-	290
Qy	156	KKILEGVVTTTSAG-----GTNNLSDSYAAGFLMLTGLMANOGIDVVIHRSFFDHGYNH	211
Dd	291	---WASPVGESGAYNSGGRHYSDFTIDSEFWLDQLGMSARHTKVYCROTLVGGFYGL	347
Qy	212	LVDQNENPLDPYULUSLYKLRIQPKULAVHVAGLQRKPGRPVIRDKLRIYAHCTNNHHH	271
Dd	348	LEKTEFPNPDPYEGALLWHRLMGKYLAVOTDG-----PPOLRVYAHCSKG---	393
Qy	272	NVVRGSTLTFTINLH-----RSRKIKLAGTURDKLV-----	303
Dd	394	--RAGVTLILLINLSNQSDFTVSVSNGINVNLNAESRKKKSLLDTLKRFPSWIGSKASDG	450
Qy	304	---HQYLLOPYGOEG-LKSKSVOLNGOPLVMYDDGFTLPCLKPRPLRAGRCTLVIPPTMG	358
Dd	451	YLNREEYHLTP-ENGVLRSKTWYNGSKLKPATGDIPSLPEPVLRSVSNPLNVPLSMS	508
Qy	359	FVVVKVNNALAC	370
Dd	509	FVLPNFEDASAC	520

### RESULT 3

T01953  
Hypothetical protein T2L5.6 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 21-Jan-2000  
C:Accession: T01953  
R:Geisel, C.; Smith, A.; Le, T.  
Submitted to the EMBL Data Library, October 1998  
A:Description: The sequence of A. thaliana T2L5.  
A:Reference number: Z14470  
A:Accession: T01953  
A:Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-190 <GEI>  
A:Cross-references: EMBL:AF096371; NID:g3695386; PID:g3695392  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4  
A:Introns: 36/2; 69/3  
A:Note: T2L5.6

C: Superfamily: Arabidopsis thaliana hypothetical protein T2L5.6

Query Match	7.9%	Score 155;	DB 2;	Length 190;
Best Local Similarity	24.9%;	Pred. No. 1.4e-05;		
Matches 49; Conservative	32;	Mismatches 70;	Indels 46;	Gaps 6;

Qy	201	RHSFFDHGYNHLVDQNFNPPLDPYWSLLLYKRLIGPKVLAVHVAGLQRPKPPGVEVIRDKLR	260
Db	12	QSLIGNGYLLNTTFTNPDPYYSALINRQLMGRKALFTTFSGTK-----KIR	60
Qy	261	IYAHCTHHHNNHNVVRSITLFIINLHRSR--KKIKLAGTLDKLVHQVLLQPY-----	311
Db	61	SYTHCARQSK-----GITVLLMNLTNTTVVAKVELNNSF--SLRHTKHKMSYKRASSQ	112
Qy	312	---GQEG-----LKRSQVQLNGPVLVMDDGTTLPELKPRLPRLAGRTLVP	353
Db	113	LFGGPGVQIOREEVHLTAKDGNLHSQTMLLNGNALQVNSMGDLPPIETPIHINSTEPITIA	172
Qy	354	PVTMGFFVVKNNVALAC	370
Db	173	PYSIFVFMNRNVVVPAC	189

## RESULT 4

reverse transcriptase homolog - slime mold (*Dictyostellium discoideum*) retrotransposab  
Tl4022  
C:Species: Dictyostellium discoideum  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 11-May-2000  
C:Accession: Tl4022  
R:Winckler, T.; Tschepke, C.; de Hostos, E.L.; Jendretzke, A.; Dingermann, T.  
Mol. Gen. Genet. 257, 655-661, 1998  
A:Title: rdd-3, a transfer RNA gene-associated poly(A) retrotransposon from Dictyoste  
A:Reference number: 217858; MUID:98265925  
A:Accession: Tl4022  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-1260 <WIN>  
A:Cross-references: EMBL:AF002669; NID:q2558820; PID:q2558822; PIDN:AAC48324.1

Query Match	5.2%	Score 102.5;	DB 2;	Length 1260;
Best Local Similarity	20.3%	Pred. No. 5;		
Matches	86:	Conservative	72:	Mismatches 160;
				Indels 105;
				Gaps 22;

[illegible]



F:243/Binding site: oxygen (Tyr) #status predicted  
F:367/Binding site: magnesium (His) (shared with chain II) #status predicted  
F:375/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 4.8%; Score 95; DB 2; Length 512;  
Best Local Similarity 24.6%; Pred. No. 6;  
Matches 42; Conservative 22; Mismatches 57; Indels 50; Gaps 11;  
QY 112 VTW-QHCYIDGRVVKVWDFLKTRELLDLSQIRKIQKVVNTYTPGKKIWLGEVVTTSAGG 170  
DB 285 IWAHMETVG-----MD-IDTRAYFAATMLIAVPTGKIFS-----WL-----ATLHGS 329  
QY 171 TNNLSDS--YAAGFLWLNTLG-----MLANQIDVIRHSFFDHGYNHLV----- 213  
DB 330 NINFSSMLWLVGFVFLPTLGLTGILANSSIDILLHDTYVVAHFHYVLSMGAVFAIM 389  
QY 214 -----DQNFNPLPDYWLSL-LYKRLIGPKV--LAVHVAGLQKPR 250  
DB 390 GSITHWFLPFGMNFSL---WLKIQFYSMFIGNMTFFPQHFLGLSSMPR 437

RESULT 8  
G90554  
p46-like (mycoplasma hyorhinis) lipoprotein [imported] - Mycoplasma pulmonis (strain UAB  
C:Species: Mycoplasma pulmonis  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001  
C:Accession: G90554  
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallissou, F.; Moszer, I.;  
Nucleic Acids Res. 29, 2145-2153, 2001  
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm  
A:Reference number: A99512; MUID:21267165; PMID:11353084  
A:Accession: G90554  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-460 <KUR>  
A:Cross-references: GB:AL445566; PID:g14089757; PIDN:CAC13516.1; GSPDB:GN00153  
A:Experimental source: strain UAB CTIP  
C:Genetics:  
A:Gene: MYP0\_3430  
A:Genetic code: SGC3

Query Match 4.8%; Score 94.5; DB 2; Length 460;  
Best Local Similarity 19.6%; Pred. No. 5, 7;  
Matches 78; Conservative 53; Mismatches 145; Indels 121; Gaps 18;  
QY 28 LKEQFSNTYSNLITLTPNNYRTMHGRAVNSQLGKDYIOLKSLLOPIRIYSRASLYGPN 87  
DB 55 VSEKISDGSIVITVDPENQ-----WVETKKQLD---AYSKT----- 89  
QY 88 GRPKNVIALLDGFMKVGASTVDVAVTWHQCYIDGRVVKVWDFLKTRELLDLSQ----- 141  
DB 90 -----NGFEHISNHVKAQAEQNSFVDAELAKTGNKPKVVLMLGAADSGNATQA 138  
QY 142 -----IRRIQK-----VNTYTPGKKIWLGEVVTTS-----GGTNKL 174  
DB 139 IESTNAAQOIAVDRFHKKISNNYVAFNNTYVGQ---LQGLALISGIYCKQGEPEK 195  
QY 175 SDSYAGFLWLNTL-----GMLANQIT-DVIRHSFFDHGYNHL-----VDQNFNPLPDYW 224  
DB 196 TLEEAKTYVANKLASEKGFVALAGAPEDNNSHLFKFGAMDVITAMKIDSNLKYFGE 255  
QY 225 LSLYKRLIGPKVLAVHVAGLQKPRGRV---IRDKLRIYACTNNHNNYVRGS---I 278  
DB 256 GADVFNLSG-----DSKPTDQELTSAITSHFRVVAQ-----WNWYSGQKQLD 300  
QY 279 TLFITNLHRSRKKIKLAGTL--RDKLVHQLQPYGOEGLKSKSVOLGQPLVWVDGTL 336  
DB 301 TLPKSTFDSRKN-NLAAVLAPNDEMAQAATISIEQGLDPKKIYITGQ-----DSNQ 353  
QY 337 PELKPRPLRAGRTLIVP-----PVTMGFFVVKV 364

Db 354 PSLSIDNETGQNMITSKEDWKIAAISTALAYYIVKN 390

RESULT 9  
G86855  
glucose inhibited division protein Gida [imported] - Lactococcus lactis subsp. lactis  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C:Accession: G86855  
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Eh  
Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis  
A:Reference number: A86625; MUID:21235186; PMID:11337471  
A:Accession: G86855  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-625 <STO>  
A:Cross-references: GB:AE005176; PID:g12724877; PIDN:AAK05945.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: gida  
C:Superfamily: gida protein

Query Match 4.8%; Score 94; DB 2; Length 625;  
Best Local Similarity 19.6%; Pred. No. 9, 6;  
Matches 72; Conservative 58; Mismatches 119; Indels 118; Gaps 17;  
QY 68 KSLLOPIRIYSRASL-YCPNIGRPKKNVIAL-LDGFPMKVAGSTVDVAVTWHQCYID---GR 122  
DB 31 KTLMTINLNVAFMPCNPISGGSAKGIVVREIDALGEGMGRNIDKTYIQMKMLNTGKGP 90  
QY 123 VVKVW-----DFLKTRELLDLSQ-----IRKI-----QKVVNTYTPGKKIWL 160  
DB 91 AVRALRAQADKDEYADSKMKNVTSQENLTLRQGMVEELILDEEKKVIGIKTSTGTYGA 150  
QY 161 EGVVTT-----SAGGTNNLSDSVAAGFLWLNTLGLMLANQIDVIRHSF 204  
DB 151 KAVIITTTGALRGIIIGELKYSNGPNNSLS-----SIGLADN-----LREIG 193  
QY 205 FDHGVNHLVDQNFNPLPDYWLSLLYKRLIGPKVLAVHVAGLQKPRPG----- 252  
DB 194 FEIG-----REFKTGTPPRVLASSIDYDKTEIQPGDEAPNHFSEWSS 234  
QY 253 --RVIRDKLRIYACTNNHNNYVRGSITLFIINLHRS---RKKIKLAG-----TLRDKL 302  
DB 235 DENVLKQIPCWLTATVTENSHILRG-----NLHRAPLFSGIVKGVGPRYCPSIDKI 287  
QY 303 V-----HQYLLQPYGOEGLKSKSVOLGQPLVWVDGTLPELKPRLRAGRTLIVIPPV 355  
DB 288 TRFADKPRHQLFLEP---EGRNTEEVYIGGLSTSPEDVDQFDLVKSIPLGNAQMMPRGY 344  
QY 356 TMGFFVW 362  
DB 345 ALEYDVV 351

RESULT 10  
B81393  
DNA-directed RNA polymerase (EC 2.7.7.6) beta' chain Cj10479 [imported] - Campylobacte  
C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 08-Sep-2000  
C:Accession: B81393  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals  
A:Reference number: A81250; MUID:20150912  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1517 <PAR>  
A:Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB75117.1; PID:g696



A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: rpoC; Cj0479

C:Superfamily: Escherichia coli DNA-directed RNA polymerase beta' chain

C:Keywords: nucleotidyltransferase

Query Match 4.8%; Score 94; DB 2; Length 1517;  
Best Local Similarity 20.2%; Pred. No. 34;  
Matches 68; Conservative 59; Mismatches 99; Indels 110; Gaps 20;

Qy 91 RKNVIALLD-----GFKVAGSTVDVNT---WQHCYIDGRVVKVMDFL----KPLDLDTL 138  
Db 607 KKDIAALVDYVYKGGLEITAFSLDRKLNGLFEYATKAGISISADIIVPNDKQKAIDEA 666  
Qy 139 SDQIRKIOKVYNTYTPGKIKWLEGVVTTSAGGTNNLSDSYAGFTLWTLGLMANOGIDV 198  
Db 667 KKQVREIQ---NSYNLG-----LITSGERYNKIID-----IWKSTNNVLSKEMKML 709  
Qy 199 VIRHSFFDHGYNHLVDQNFNLPDYWLSLLYKRLIGPKVLAVHVA-----GLQRKPRPG 252  
Db 710 VEKDK---EGFN-----SIYMMADSGARGSAQISQLAAMRGLTKP-DG 750  
Qy 253 RVIRDKLRIYAHCTNHHNNHVVG-SITFLFIINLHRSRK-----KIKLAGTLRDKLHV 304  
Db 751 SIETPI-----ISNREGNLNVLEYFISTHGARKGLADTALKTANAGAYLTRKLIID 800  
Qy 305 -----QYLLOPYGOGLKSKSVOLNGQPLVMVDGTLPELKPRLRAGRTL---VIPPV 356  
Db 801 VAQNVKITIEDCGTH---EGVEINE---ITADSSIIETLEERIL--GRVLAEDVIDPIT 851  
Qy 357 -----MGFFVKNVN---ALACR 371  
Db 852 NSVLFAEGLTMDDEKAKILGESGIKSVNIRTPITCK 887

RESULT 11

D90478

conserved hypothetical protein [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001

C:Accession: D90478

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, B.  
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: D90478

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-320 <KUR>

A:Cross-references: GB:AE006641; NID:g13816361; PIDN:AAK43083.1; GSPDB:GN00155

C:Genetics:

A:Gene: SS02979

Query Match 4.8%; Score 93; DB 2; Length 320;  
Best Local Similarity 23.7%; Pred. No. 4.6;  
Matches 56; Conservative 31; Mismatches 77; Indels 72; Gaps 12;

Qy 23 MHLVLLKEQFSNTYSNLILTEPNNTYTMHGRAVNGSQLGKDYIOLKSLLOPIRIYSRASL 82  
Db 1 MNLATIEFE-SNVLDRNPLNDPYKRRV-----GIIPKDYEG-----RPILVLSGYL 47  
Qy 83 -----YGP-----NIGPRKNVIALLDGFKMKVAGSTVDVAVTQHCYIDG 121  
Db 48 SSSLTQIINPLGEDMWSKVERLSNEGKMGSVIPLDMFTKVGQ-----YINS 98  
Qy 122 RVVKVW-DFLKRLDLDLSQIRKIQKVVNTYTPGKIKWLEGVVTTSAGGTNNLSDSYAA 180  
Db 99 SAVGWYEDFLVKELIPFLKDF-KSDKI-----GIFCHSSGGYGAL-----IL 140

Qy 181 GFLWLNTGLMANOGIDVIRHSFFDHGYNHLVDQNFNPL-----PDYWLSLLYKR 231  
Db 141 GMKYPNTIKAIADHAGD-----AVFEYVYLTPFPRAIEQLRRFKTPEEWLENYWKK 191

RESULT 12

AC1210

dTDP-D-glucose 4,6-dehydratase homolog lmo1083 [imported] - Listeria monocytogenes (s

C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C:Accession: AC1210

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,  
D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Krefet, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;  
Ok, C.; Schluter, T.; Smoes, N.; Tierrez, A.; Vasquez-Boland, J.A.; Voss, H.; Wehla

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AC1210

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-328 <GLA>

A:Cross-references: GB:NC\_003210; PIDN:CAC99161.1; PID:g16410485; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lmo1083

C:Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase homolo

Query Match 4.7%; Score 92.5; DB 2; Length 328;  
Best Local Similarity 20.4%; Pred. No. 5.2;  
Matches 57; Conservative 41; Mismatches 91; Indels 91; Gaps 15;

Qy 116 HCYIDGRVVKVMDFLKRLDLDLSQIRKIQKVVN-----TYTPGKKIWLEGVVT 165

Db 53 HVFVEGNICD-YDLVKKLVTD-----HKIDTIVFAAESHVDRSIINPG---IFIE---- 99

Qy 166 TSAGTNNL-----SDSYAAGFLWNT---LGLMANOG-----IDV 198

Db 100 TNVQGTLLNNVAKELNVAKYLVQVTSDEYVSLGTFYFTETPIAPNSPYSKASADL 159

Qy 199 VIRHSFFDHGYNHLV---DQFNPP---LPDYWLSLLYKRLIGPKVLAVHVGLOKRPGR 253

Db 160 LVRSYFYETGLNVNITRCSNNYPHPFKLPLMITNGDGENLPY-----GDGK 211

Qy 254 VIRDKLRIYANC-----TNNHHNNYVRSITFLIINLHRSRKKIKLAG 296

Db 212 NIROWLHVSDHCAADLVINHCKSGEVYVNGHNERTNNEIHHIIVDDLNLJSKDKIVY-- 269

Qy 297 TLIRDKLHV---QYLLOPYGOGLKSKSVOLNGQPLVMVDG 334

Db 270 -VEDRLGHDLRYAIDP-----KKIETELGWEPKYTFDTG 302

RESULT 13

G89952

DNA polymerase I [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C:Accession: G89952

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; K  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: G89952

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-876 <KUR>

A:Cross-references: GB:BA000018; PID:g13701486; PIDN:BAB42780.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:  
A:Gene: polA  
C:Superfamily: DNA-directed DNA polymerase I

Query Match 4.7%; Score 92.5; DB 2; Length 876;  
Best Local Similarity 21.7%; Pred. No. 21;  
Matches 62; Conservative 51; Mismatches 104; Indels 69; Gaps 14;

Qy 22 QHVLVLLKEQFSNTYSNLI--LTPN-----NYRTMHGRAVNSQLGKDYIQLKS 69  
Db 342 EKHIIVINADDINN-YAELVSLNLEPNTKKVYDAKKTIVASHRLGIDQIONISFDIMIASY 400  
Qy 70 LQIPTR----IYSRASLYGPNI-----GPRKNVIALLDGFMKVGAGSVDAVTWQHC 117  
Db 401 IIDPRTTSDVQSVVSLYGQSFVKDDVSIYGGKKFKVPEDDLNPNYVASITDAIYFAKP 460  
Qy 118 YIDGRV-----YKVMDFLTKRLDITLSD-----QIRKIQKVNTYTPGKKIWL 161  
Db 461 NMDKQLEENQVELLADLEPLAKITLSEMEETIGFTDVHDLSEMEKEIQE-----KLDVLI 516  
Qy 162 GYVTSAGCTNNLSDSYAGFLWLTNLGLMANOGIDVIRHSFFDGHYNNHLVD-----QN 216  
Db 517 RNHDAAGDFNINPKQGVVLFETLQ-----PVIKKT--KTGYSTAVDVLQEQLOG 567  
Qy 217 FNPDPYLSLLYKRLIGPKVLAVHAGLQKRPGRVIRKRIY 262  
Db 568 EHPIDYILE--YROL--SKLQSTVVEGLQ-----KVISDDORIH 603

RESULT 14  
S09811  
hypothetical protein UL48 - human cytomegalovirus (strain AD169)  
N:Alternate names: hypothetical protein HPRF0  
C:Species: human cytomegalovirus, human herpesvirus 5  
A:Note: host Homo sapiens (man)  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 08-Oct-1999  
C:Accession: S09811  
R:Chee, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Horsnell, T.; M.; Barrell, B.G.  
Curr. Top. Microbiol. Immunol. 154, 125-169, 1990  
A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus  
A:Reference number: S09749; MUID:90269039  
A:Accession: S09811  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2241 <CHE>  
A:Cross-references: EMBL:X17403; NID:g959591; PIDN:CAA35407.1; PID:g1780826  
A:Note: this sequence was submitted to the EMBL Data Library, December 1989

Query Match 4.7%; Score 92.5; DB 2; Length 2241;  
Best Local Similarity 20.8%; Pred. No. 78;  
Matches 78; Conservative 54; Mismatches 110; Indels 133; Gaps 21;

Qy 15 LQREKAQMHVLVLKQFNTYSNLIITPE-----NNYRTMHGRAVNSQLGKDYIQ 66  
Db 1325 LSAERTVWMLAFLEELAQ-----ITADTPHGHGSRTHYRNLLQQAQVESAVTLAQHIE 1379  
Qy 67 LKSLQIP--IRIYSRASLYGPNIGRKNVIALLDGFMKVGAGSVDAVTWQHCVIDGRVV- 124  
Db 1380 QNAACENFIAGHQEATANGAS--TPRVDV-----QAVEAV-WORLE-PGRVAG 1424  
Qy 125 -----KVMDFLTKRLDITLSD-----QIRKIQKV 149  
Db 1425 GAARHQKVOELLQ--RLGQTGLDELQETLATEVFALLHGQTFSYGLDFFRSQLEKINDLR 1483  
Qy 150 NTYTP-----GKKIWLQGVV-----TTSAGCTNNLSDSYAGFLWLTNLGLMANQGI 196  
Db 1484 TRFAELAKRRGTRLSNKGVLNPRKPQATISLGA-----FTRGL 1522  
Qy 197 DVIYHHSFFDHCY--NHLVDQNFNLPDYWLSLYLKRLLIGPKVL-AVHVAGLQKRPGR 253

Db 1523 NALERHVQLGHOYLNLKNGS-----SLVYRLEDIPSVLPATH-----ETDPAL 1566  
Qy 254 VIRDKLRIYAHCTNNHHNNHNVYRGSTLFTIINLHRSKKIKLAGTLRDLKLVHQYLLQPYGQ 313  
Db 1567 IMRDRLR--RLCFARHDTF-----LEVVDVFGMRQIVTQAG-----EPIH--LVTDYGN 1612  
Qy 314 EGLKSKSVQLNGOPL 328  
Db 1613 VAFKYLALRDDGRPL 1627

RESULT 15  
S65187  
GPI-anchor biosynthesis protein PIG-A - yeast (Saccharomyces cerevisiae)  
N:Alternate names: GlcNAc-inositol phospholipid assembly protein; GPI-anchor biosynthesis  
C:Species: Saccharomyces cerevisiae  
C:Date: 10-Dec-1994 #sequence\_revision 31-May-1996 #text\_change 16-Jun-2000  
C:Accession: S65187; S27131; A61112; S18842; S60357  
R:Benes, V.; Rechmann, S.; Nentwich, U.; Voss, H.; Ansoerge, W.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S65183  
A:Accession: S65187  
A:Molecule type: DNA  
A:Residues: 1-461 <BEN>  
A:Cross-references: EMBL:Z73531; NID:gl370368; PIDN:CAA97882.1; PID:gl370369; GSPDB:G  
A:Experimental source: strain S288C (AB972)  
R:Fassler, J.  
submitted to the EMBL Data Library, January 1992  
A:Reference number: S27131  
A:Accession: S27131  
A:Molecule type: DNA  
A:Residues: 16-461 <FAS1>  
A:Cross-references: EMBL:X63290  
A:Accession: A61112  
A:Molecule type: DNA  
A:Residues: 'MGF', 'II', 'IA', '16-461' <BAU>  
A:Cross-references: EMBL:X63290; NID:g4533; PIDN:CAA44924.1; PID:g441475  
R:Fassler, J.S.; Gray, W.; Lee, J.P.; Yu, G.; Gingerich, G.  
Mol. Gen. Genet. 230, 310-320, 1991  
A:Title: The Saccharomyces cerevisiae SPT14 gene is essential for normal expression o  
A:Reference number: S18842; MUID:92079912  
A:Accession: S18842  
A:Molecule type: DNA  
A:Residues: 16-429, 'S' <PAS2>  
A:Cross-references: EMBL:X63290  
R:Vossen, J.H.; Ram, A.F.J.; Klis, F.M.  
Biochim. Biophys. Acta 1243, 549-551, 1995  
A:Title: Identification of SPT14/CWH6 as the yeast homologue of hPIG-A, a gene involv  
A:Reference number: S60357; MUID:95244613  
A:Accession: S60357  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 16-461 <VOS>  
C:Genetics:  
A:Gene: SGD:SPT14; CWH6; GPI3; MIPS:YPL175W  
A:Cross-references: SGD:S0006096; MIPS:YPL175W  
A:Map position: 16L  
A:Introns: 5/2  
C:Superfamily: GPI-anchor biosynthesis protein PIG-A  
C:Keywords: glycoprotein; transmembrane protein  
F:416-434/Domain: transmembrane #status predicted <TMM>  
F:140,386/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 4.7%; Score 92; DB 1; Length 461;  
Best Local Similarity 22.0%; Pred. No. 9.2;  
Matches 50; Conservative 38; Mismatches 63; Indels 76; Gaps 12;  
Qy 181 GFLLWLTGLMLANQGDVIRHSFEDH---GYNHLVDQNFNLPDYWL-----225  
Db 117 GILHANTMG-----LRTVFTDHSLYGFNNLTSIWNKLLFTLTNIDRVICVSN 166  
Qy 226 ---SLLYKRLIGPKVLAV---HVAGLQKRP-----PGRVTRDKL-----RIYAH--- 264

```
Db 167 CKENHIVRTELSPDIIISVIPNAVVSDEFKPRDPTGGTKRKQSRDKIIVIVIGRLFPNKG 226
Qy 265 -----CTNHHNNY-VRGSTITLFI-----INLHRSRKIKLACT-----LRDKL- 302
Db 227 DLLTRIIPKVCSSHEDVEFIVAGDGPKEFIDFQQMIESHRLQKRVQLLGSVPHEKVRDVL 286
Qy 303 -----VHQYLLQPYQEGGLKSKSVQLNGOPLVMVDDGTLPCLKPRPL 344
Db 287 QGDIYLHASLTEAFGTILVEAASCNL---LIVTTQVGGIPEVLPNEM 330
```

Search completed: July 30, 2002, 08:16:34  
Job time: 351 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:31:33 ; Search time 36.14 Seconds  
(without alignments)  
399.623 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_162\_534  
Perfect score: 1956  
Sequence: 1 KGCKTAQHPDVMLEQLREKA.....PVTMGFFVKNYNALACRYR 373

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100.5	5.1	630	1	GIDA_TREPA
2	94	4.8	625	1	GIDA_LACLA
3	94	4.8	1517	1	RPOC_CAMJE
4	92.5	4.7	2241	1	TEGU_HCMVA
5	92	4.7	461	1	GP13_YEAST
6	92	4.7	534	1	COX1_KLULA
7	92	4.7	535	1	COX1_HANWI
8	91.5	4.7	335	1	SYW_BUCAT
9	90	4.6	1616	1	VITL_CAEEL
10	88.5	4.5	286	1	PARB_CHLPN
11	88.5	4.5	341	1	YF43_MYCTU
12	88	4.5	1106	1	CIC2_SABIT
13	87.5	4.5	534	1	COX1_SACDO
14	87.5	4.5	557	1	COX1_NEUCR
15	87.5	4.5	627	1	MUTL_BAGSU
16	87.5	4.5	828	1	YF44_YEAST
17	87	4.4	528	1	NIFK_FRAAL
18	87	4.4	535	1	GIDA_LACIC
19	85.5	4.4	512	1	HEXB_YEAST
20	85.5	4.4	773	1	COX1_ALTZO
21	85.5	4.4	898	1	ACOC_CUCMA
22	84.5	4.3	552	1	DP1B_ECOLI
23	84.5	4.3	434	1	TI47_HUMAN
24	84.5	4.3	928	1	CN2A_RAT
25	84	4.3	514	1	COX1_PROMI
26	84	4.3	621	1	GIDA_RICPR
27	84	4.3	905	1	COPG_SCHPO
28	83	4.2	401	1	ACHN_CHICK
29	83	4.2	536	1	DIT1_YEAST
30	83	4.2	838	1	SVV_MYCPN
31	83	4.2	934	1	SVL1_SULSO
32	82.5	4.2	527	1	COX1_OENBE
33	82.5	4.2	541	1	COX1_PODAN

34	82.5	4.2	564	1	DNL1_FOWPV
35	82.5	4.2	921	1	CN2A_BOVIN
36	82.5	4.2	1065	1	SP18_SCHPO
37	82.5	4.2	1118	1	VIPL_YEAST
38	82	4.2	527	1	COX1_SOYBN
39	82	4.2	591	1	PAK4_HUMAN
40	82	4.2	854	1	AKA3_BOVIN
41	82	4.2	1399	1	RPOC_PSEAE
42	82	4.2	1541	1	MRP2_RAT
43	81.5	4.2	477	1	EXU2_DROPS
44	81.5	4.2	645	1	LIP1_PHOLU
45	81.5	4.2	687	1	TGLC_BOVIN

ALIGNMENTS

RESULT 1					
GIDA_TREPA					
ID	GIDA_TREPA	STANDARD;	PRT;	630 AA.	
AC	083084;				
DT	15-DEC-1998 (Rel. 37, Created)				
DT	15-DEC-1998 (Rel. 37, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Glucose inhibited division protein A.				
GN	GIDA OR TP0044.				
OS	Treponema pallidum.				
OC	Bacteria; Spirochaetales; Spirochaetaceae; Treponema.				
OX	NCBI_TaxID=160;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=NICHOLS;				
RX	MEDLINE=98332770; PubMed=9665876;				
RA	Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,				
RA	Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,				
RA	Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,				
RA	Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,				
RA	McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,				
RA	Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,				
RA	Venter J.C.;				
RT	"Complete genome sequence of Treponema pallidum, the syphilis				
RT	spirochete.";				
RL	Science 281:375-388(1998).				
CC	-1- FUNCTION: NOT KNOWN.				
CC	-1- SIMILARITY: BELONGS TO THE GIDA FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; AE001189; AAC65038.1;				
DR	TIGR; TP0044;				
DR	InterPro; IPR002218; GIDA.				
DR	Pfam; PF01134; GIDA; 1.				
DR	ProDom; PD003738; GIDA; 1.				
DR	PROSITE; PS01280; GIDA_1; 1.				
DR	PROSITE; PS01281; GIDA_2; 1.				
KW	Complete proteome.				
SQ	SEQUENCE 630 AA; 70175 MW; 1B52G6F02C1EC275 CRC64;				

Query Match 5.1%; Score 100.5; DB 1; Length 630;  
Best Local Similarity 21.3%; Pred. No. 1.1;  
Matches 77; Conservative 52; Mismatches 129; Indels 103; Gaps 20;

Qy	57	GSOLGK--DYIOLKSLQPIRIYSRASLYGNIGRPNKVNIALLDGFMKVAGSTVDVATW	114
Db	68	GEEMGKFDACMIQ-----YRLNLSR--GPAVQAPRIQADKFL--YAKVKVYLTCTQH	118





RX MEDLINE=95255221; PubMed=7737116;  
 RA Schoenbacher M., Horvath A., Fassler J.S., Riezman H.;  
 RT "The yeast spt14 gene is homologous to the human PIG-A gene and is  
 RL required for GPI anchor synthesis.";  
 CC EMBO J. 14:1637-1645(1995).  
 CC -1- FUNCTION: NECESSARY FOR THE SYNTHESIS OF N-ACETYLGALACTOSAMINYL-  
 CC BIOSYNTHESIS.  
 CC -1- PHOSPHATIDYLINOSITOL, THE VERY EARLY INTERMEDIATE IN GPI-ANCHOR  
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.  
 CC -1- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE INVOLVED IN  
 CC TRANSCRIPTION.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; X63290; CAA44924.1; -;  
 CC EMBL; Z73531; CAA97882.1; -;  
 CC PIR; S27131; S27131.  
 CC PIR; S18842; S18842.  
 CC SGD; S0006096; SPT14.  
 CC InterPro; IPR001296; Glycos\_transf\_1.  
 CC Pfam; PF00534; Glycos\_transf\_1; 1.  
 CC Transferase; Glycosyltransferase.  
 CC CONFLICT 1 15 MSSSHKVESYNENR -> MGFNIA (IN REF. 1).  
 CC SEQUENCE 461 AA; 52372 MW; 3DDC44DD5DA43E5A CRC64;  
 CC -----

Query Match 4.7%; Score 92; DB 1; Length 461;  
 Best Local Similarity 22.08; Pred. No. 3.6;  
 Matches 50; Conservative 38; Mismatches 63; Indels 76; Gaps 12;  
 Qy 181 GFLMNTLGLMANGIDVIRHSFDH---GYNHLVDQNFNPLPDYWL----- 225  
 Db 117 GILHMTG-----LRTVETDLSLGFNLTISWKNLLTFTLTNIDRVICVSNT 166  
 Qy 226 ---SLLYKRLGPKVLAV---HVAGLQRPKPR-----PGRVIRDKL-----RIYAH--- 264  
 Db 167 KENMIVTELSPLDIISIPNAVSEDFKPRDPTGGTKRKOSDKIVIVIGRIFPNKGS 226  
 Qy 265 -----CTNNHNNHY-VRGSTITLFI-----INLHRSRKIKLAGT-----LRDKL- 302  
 Db 227 DLLTRIIPKVCSSHEDVEFVAGDGPKEFDQMTESHRLQKRVQLLGSVPHEKVRDVL 286  
 Qy 303 -----VHQYLQPGQEGLSKSVOLNGQPLVMVDDGTLPKPRPL 344  
 Db 287 QGDIYLHASLTFEATGILVEAASCNL---LIVTTQVGGIPEVLPNEM 330

RESULT 6  
 COX1\_KLULA  
 ID COX1\_KLULA STANDARD; PRT; 534 AA.  
 AC P20386;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).  
 GN COX1.  
 OS Kluyveromyces lactis (Yeast).  
 OC Mitochondrion.  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OX NCBI\_TaxID=28985;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K8;  
 RX MEDLINE=92035081; PubMed=1657415;  
 RA Hardy C.M., Clark-Walker G.D.;  
 RT "Nucleotide sequence of the COX1 gene in Kluyveromyces lactis

RT mitochondrial DNA: evidence for recent horizontal transfer of a group  
 RT II intron.";  
 RT Curr. Genet. 20:99-114(1991).  
 RN [2]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RC STRAIN=K8;  
 RX MEDLINE=91021490; PubMed=2171241;  
 RA Hardy C.M., Clark-Walker G.D.;  
 RT "Nucleotide sequence of the cytochrome oxidase subunit 2 and val-crna  
 RT genes and surrounding sequences from Kluyveromyces lactis K8  
 RT mitochondrial DNA.";  
 CC Yeast 6:403-410(1990).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
 CC AND COPPER B.  
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome  
 CC c + 2 H(2)O.  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; X57546; CAA40765.1; -;  
 CC EMBL; X15999; CAA34130.1; -;  
 CC HSSP; P00396; IOCC.  
 CC InterPro; IPR000883; COX1.  
 CC Pfam; PF00115; COX1; 1.  
 CC PRINTS; PR01165; CYCOXIDASE1.  
 CC PROSITE; PS00077; COX1; 1.  
 CC Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;  
 CC Respiratory chain; Inner membrane.  
 CC METAL 62 62 IRON (HEME A) (PROBABLE).  
 CC METAL 241 241 COPPER B (PROBABLE).  
 CC METAL 245 245 COPPER B (PROBABLE).  
 CC METAL 290 290 COPPER B (PROBABLE).  
 CC METAL 291 291 COPPER B (PROBABLE).  
 CC METAL 376 376 IRON (HEME A3) (PROBABLE).  
 CC METAL 378 378 IRON (HEME A) (PROBABLE).  
 CC SEQUENCE 534 AA; 58677 MW; 4A26B1C323F9F1C2 CRC64;  
 CC -----

Query Match 4.7%; Score 92; DB 1; Length 534;  
 Best Local Similarity 22.98; Pred. No. 4.4;  
 Matches 72; Conservative 39; Mismatches 97; Indels 106; Gaps 19;  
 Qy 74 IRIYSRASLYG-----PNIGRPKNVIALLDGFMKVGAGSTVDVAVTW-QHCYIDGRVV 124  
 Db 259 VSTYSKRPVFGESVMYAMASIG-----LLGFL-----VNSHMIYVG--- 296  
 Qy 125 KVMDFLTKRLDITLSDQIRTKQKVNTYTPGKKIWLGEVVTTSAGGTNNLSDS--YAAGF 182  
 Db 297 --LD-ADTRAVFTSATMIIAIPGTGIKIFS-----WL-----ATIVGGSLRAVPMYATAF 344  
 Qy 183 LWLNTLIG-----MLANQGDVIRHSFDHGYNHLY---DQNFNPLPDYW-----LSLL 228  
 Db 345 LFLFTIGLGTVALANASLDVAFHDTYVVVGHFHYLSMGAIFSLFAGFYWSPQILGLY 404  
 Qy 229 YKR-----LIGPKV--LAVHVAGLQRPKPRGVRIRDKLRIYAHCTNNHNNHYVRG 276  
 Db 405 YNEKLAQTQFWLIFGANVIFLPMHFLGVNGMPR-----RIPDYPDAFAGNYYVAS 455



```
QY 277 -----SITLFIINLHRSRKKIKLAGTLRDKLVHQVYLL-----QPYGQEGLEK 317
Db 456 IGSIIAIFSLFIYLD-----QLVNGLENKVNKNSVIYNKGPDPFVESNQIFATNRIK 510
QY 318 SKSVQ--LNGOPLV 329
Db 511 SSSIEFLTSPAV 524

RESULT 7
COX1_HANWI STANDARD; PRT; 535 AA.
AC P48868;
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
GN COX1.
OS Hansenula wingei (Yeast).
OG Mitochondrion.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4907;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=21;
RA Sekito T., Okamoto K., Kitano H., Yoshida K.;
RT "Yeast Hansenula wingei mitochondria genome's complete DNA sequence
RL Nucleic Acids Symp. Ser. 31:233-234(1994).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B.
CC -1- CATALYTIC ACTIVITY: 4 ferrocycytochrome c + O(2) = 4 ferrocycytochrome
CC c + 2 H(2)O.
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D31785; BAA06563.1; -.
CC HSPSP; P00396; LOC.
CC InterPro: IPR000883; COX1.
CC Pfam: PF00115; COX1; 1.
CC PRINTS: PR01165; CYCOXIDASE1.
CC PROSITE: PS00077; COX1; 1.
KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW Respiratory chain; Inner membrane.
FT METAL 63 63 IRON (HEME A) (PROBABLE).
FT METAL 242 242 COPPER B (PROBABLE).
FT METAL 246 246 COPPER B (PROBABLE).
FT METAL 291 291 COPPER B (PROBABLE).
FT METAL 292 292 COPPER B (PROBABLE).
FT METAL 377 377 IRON (HEME A3) (PROBABLE).
FT METAL 379 379 IRON (HEME A) (PROBABLE).
SQ SEQUENCE 535 AA; 59416 MW; 5302237917443D38 CRC64;

Query Match 4.7%; Score 92; DB 1; Length 535;
Best Local Similarity 22.7%; Pred. No. 4.4;
Matches 69; Conservative 42; Mismatches 103; Indels 90; Gaps 19;
```

```
QY 74 IRIYGRASLYGNIGRPRKRNIALDGFMKVAGSTVDVTV--QHCYIDGRVVKVMDFLKT 132
Db 260 VSTYSKKPFGVGMSSVYAMASIAFL-GFL-----VWSHMYIVG-----LD-ADT 302
QY 133 RLDDTLSDQIRKIQKVVNTYTPGKKIWIIEGVVTTSGAGTNNLS--DSYAAGFLWNTLIG- 189
Db 303 RAYFTSSTMVIAVPTGKMF-----WL-----ATLYGGSIRLAVPMTVAIAFLFLTMGG 353
QY 190 ----MLANOGIDVIRHSHFEDHGYNHLV--DQNFENPLPDYV-----LSLLYKR----- 231
Db 354 LTGVALANASLDVAFHDTYYVVGHEHYVLSMGAIISLFAGIYVNSPQILGLFNERLAQM 413
QY 232 -----LIGPKV--LAVHVAGLQKRPGRVIRDKLRIYAHCTNHHNNYVRG----- 276
Db 414 QFWLIFVGANVIFMPHFGLTGTGMPR-----RMPDYPDAYAGWNVVSSMGSMVMAIM 464
QY 277 SITLFIINLHRSRKKIKLAGTLRDKLVHQ-----YLLQP-----YQEGLEKSKSVQ--L 323
Db 465 SLALFIYIMYD-----QLINGLTNRKMDNKSVVYSKAPDFVESNTMFANNSIKSASIEFLL 519
QY 324 NGQP 327
Db 520 NSPP 523

RESULT 8
SYW_BUCAI STANDARD; PRT; 335 AA.
AC PSY602;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Tryptophan--tRNA ligase)
DE (TrpRS).
GN TRPS OR BU536.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-tryptophan + tRNA(Trp) = AMP +
CC diphosphate + L-tryptophanyl-tRNA(Trp).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINONACYL-TRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP001119; BAB13229.1; -.
CC InterPro: IPR001412; tRNA-synt_1.
CC InterPro: IPR002306; tRNA-synt_trp.
CC Pfam: PF00579; tRNA-synt_1b; 1.
CC PRINTS: PS01039; TRNASYNTHTRP.
CC PROSITE: PS00178; AA_TRNA_LIGASE_1; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 14 22 "HIGH" REGION.
FT SITE 196 200 "KMSKS" REGION.
FT BINDING 199 199 ATP (BY SIMILARITY).
```

```
SQ SEQUENCE 335 AA; 38549 MW; 871562D5A6734E3F CRC64;

Query Match
Best Local Similarity 4.7%; Score 91.5; DB 1; Length 335;
Matches 46; Conservative 33; Mismatches 64; Indels 65; Gaps 10;

QY 22 OMHLVL---LKEQFSNTYSNLLITPENNYRTMHGRAVNSQLGKDYIOLKSLLOPIRIYS 78
DB 152 KQVELTRNIAHRFNSLYGH-VFTLPKPLITQHS-----KMSLLEPSKKMS 198
QY 79 RASLYGPNIGRPRKNVIALLDG---FMKVAGSTVDVATWQHCYID-----GRVVKVM 127
DB 199 KSDI-----NKNKVFLLDDIKTVISKIQNAYTDSETPSKIYYDIEKKPGISNLLLETL 251
QY 128 -----DFL-----KRLDITLSDQIRKIOKVNTYTPG-----KKWLEGV 163
DB 252 SAITNKDIDILLKEGLMYSEFNKIVADHLSKFLYKLOKSYRNDENVYLLKTIAYEGA 311
QY 164 VTTSGAGTNNLSDSYAAGFLWNTLQML 191
DB 312 MKSQLSKNTLTKVY-----DKLGLI 332

RESULT 9
VIT1_CAEEL STANDARD; PRT; 1616 AA.
AC P55155;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Vitellogenin 1 precursor.
GN VIT-1 OR K09F5.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Bentley D.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PRECURSOR OF THE EGG-YOLK PROTEINS THAT ARE SOURCES OF
CC NUTRIENTS DURING EMBRYONIC DEVELOPMENT (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U37430; AAB52675.1; -.
DR WormPep; K09F5.2; CE04746.
DR InterPro; IPR001747; Vitellogenin_N.
DR InterPro; IPR001846; Vwd.
DR Pfam; PF01347; Vitellogenin_N; 1.
DR Pfam; PF00094; vwd; 1.
DR SMART; SM00216; VWD; 1.
KW Storage protein; Multigene family; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1616 VITELLOGENIN 1.
SQ SEQUENCE 1616 AA; 188063 MW; F173D7C452D123F5 CRC64;

Query Match
Best Local Similarity 4.6%; Score 90; DB 1; Length 1616;
Matches 76; Conservative 68; Mismatches 131; Indels 96; Gaps 17;

QY 26 VLKKEQFSNTY---SNLLITPENNYRTMHGRAVNSQLGKDYIOLKSLLOPIRIYSRASLY 83
DB 233 VLIRPQTVTYIILENELKE-SEVRSLYTVNVANGQEVMTETRSKLVLE--ENHSIKS-H 288
```

```
QY 84 GPNIGRPRKNVI-----ALLDGEKMKVAGSTVDVATWQHCYIDGRVVKVMDFLKTRLDDT 137
DB 289 IEKVNGEKESIYSRWELQVDEDFK-NGDKAEFAFPFKFPLD-----KKMHLTKT 338
QY 138 LSDQIRKIQKVNTYTPGKKIWLEGVV---TTSAGTNNLSDSYAAGFLWNTLIGMLAN 193
DB 339 ITEQIOEVENNI---PETSHPFLARLVRIFRTTS---TSQKEIH-----ETLYVKAD 384
QY 194 QGIDVVRHSF-----FDHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKVLAVHVAGL 245
DB 385 KKIQSLMEHALAIAIACTKNTIOHILVHIENEDIVPLE-----AAQLKSI 428
QY 246 QRKPPGRVIRDKLRIYAHCTNHHNNYVR-----GSITLFIINLHRSR----- 289
DB 429 QETPFSQTIAEALIKFAESRVSKNNQVRSQAWLAAGSVVRGIVDYKNIRFLVREDKRE 488
QY 290 -----KKIKIAGTLRDKLVHOYLLQPYQGQGLSKSVQNLNGQPLVWVDDGTLPDL 339
DB 489 LKEKELRVFMQYKDAETTYEKIL---ALKSIGNAGLDISVNQLNE---IIVDKRQLLPV 542
QY 340 KPRPLRAGRTL 350
DB 543 RKEAIDALRLL 553

RESULT 10
PARB_CHLPN STANDARD; PRT; 286 AA.
AC Q9Z7M0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Probable chromosome partitioning protein parB.
GN PARB OR CPN0684 OR CP0062.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heldelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -!- FUNCTION: INVOLVED IN CHROMOSOME PARTITION. LOCALIZE TO BOTH POLES
CC OF THE PREDIVISIONAL CELL FOLLOWING COMPLETION OF DNA REPLICATION.
CC BINDS TO THE DNA ORIGIN OF REPLICATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE PARB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```



OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88336904; PubMed=2458626;  
 RA Ellis S.B., Williams M.E., Ways N.R., Brenner R., Sharp A.H.,  
 Leung A.T., Campbell K.P., McKenna E., Koch W.J., Hui A.,  
 Schwartz A., Harpold M.M.;  
 RT "Sequence and expression of mRNAs encoding the alpha 1 and alpha 2  
 subunits of a DHP-sensitive calcium channel.";  
 RL Science 241:1661-1664(1988).  
 RN [2]  
 RP SEQUENCE OF 961-973.  
 RX MEDLINE=91131638; PubMed=1847144;  
 RA Jay S.D., Sharp A.H., Kahl S.D., Vedvick T.S., Harpold M.M.,  
 Campbell K.P.;  
 RT "Structural characterization of the dihydropyridine-sensitive calcium  
 channel alpha 2-subunit and the associated delta peptides.";  
 RL J. Biol. Chem. 266:3287-3293(1991).  
 RN [3]  
 RP SEQUENCE OF 961-975; 992-1000 AND 1033-1050.  
 RX MEDLINE=90368635; PubMed=2168391;  
 RA de Jongh K.S., Warner C., Catterall W.A.;  
 RT "Subunits of purified calcium channels. Alpha 2 and delta are encoded  
 by the same gene.";  
 RL J. Biol. Chem. 265:14738-14741(1990).  
 CC -!- FUNCTION: CALCIUM CHANNEL PROTEIN WHICH PLAYS AN IMPORTANT ROLE IN  
 CC EXCITATION-CONTRACTION COUPLING.  
 CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:  
 CC ALPHA-1, ALPHA-2, BETA AND GAMMA. ALPHA-2 AND DELTA FORMS  
 CC HETERODIMERS THAT ARE DISULFIDE-LINKED.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS ARE PRODUCED BY ALTERNATIVE  
 CC SPLICING.  
 CC -!- TISSUE SPECIFICITY: SKELETAL MUSCLE.  
 CC -!- PTM: ALPHA-2 AND DELTA SUBUNITS ARE PROTEOLYTICALLY PROCESSED FROM  
 CC A PRECURSOR FORM.  
 CC -!- SIMILARITY: TO OTHERS SPECIES ALPHA-2 SUBUNIT.  
 CC -!- SIMILARITY: CONTAINS 1 VFMA DOMAIN.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M21948; AA81562.1; -;  
 DR PIR; S10579; CHRB2.  
 DR InterPro; IPR004010; Cache.  
 DR InterPro; IPR002035; VFMA.  
 DR Pfam; PF02743; Cache; 1.  
 DR Pfam; PF00092; vwa; 1.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS0234; VFMA; 1.  
 KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;  
 KW Calcium channel; Glycoprotein; Phosphorylation; Signal;  
 KW Alternative splicing.  
 FT SIGNAL 1 26  
 FT CHAIN 27 960  
 FT CHAIN 961 1106  
 FT TRANSMEM 448 471  
 FT TRANSMEM 921 945  
 FT TRANSMEM 1082 1101  
 FT DOMAIN 255 432  
 FT CARBOHYD 94 94  
 FT CARBOHYD 138 138  
 FT CARBOHYD 186 186  
 FT CARBOHYD 326 326  
 FT CARBOHYD 350 350  
 FT CARBOHYD 477 477  
 FT CARBOHYD 606 606

FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 678 678 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 784 784 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 891 891 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 898 898 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 988 988 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1001 1001 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT MOD\_RES 503 503 PHOSPHORYLATION (BY CAPK)  
 FT MOD\_RES 848 848 PHOSPHORYLATION (BY CAPK)  
 FT SEQUENCE 1106 AA; 125042 MW; B00DE7F3C877B618 CRC64;  
 SQ  
 Query Match 4.5%; Score 88; DB 1; Length 1106;  
 Best Local Similarity 18.7%; Pred. No. 24; Indels 122; Gaps 14;  
 Matches 68; Conservative 55; Mismatches 119;  
 QY 92 KNVIALDGFEMKVGSTVDVAVTWQHCYIDGRVVKVMDFLKRLDLTLD----- 140  
 DB 254 KDMILVD-----VSGS-----VSLTLKLRISVSSEMLETSLDSDDFNVASFNS 298  
 QY 141 ---QIRKIQKVVNTYTPGKIMLEGVVTTSGAGTNNLSDSYAAGFLWL-----NT 187  
 DB 299 NAQDVSCFQHLVQAVNRNKKVLKDAVNNITAKGITYDKGFSFAFQOLLNVNVRANCK 358  
 QY 188 LGMLANOGIDVIRHSEDFHGYNHLV-----DQNFNPLPDYWLSL----- 227  
 DB 359 IIMFTDGGERAQEIFAKYNKKDKVRFVTSVQHNVDGRPIOMACENKGYVYEIPI 418  
 QY 228 -----LYKRLIG-PKVL-----VHVAGLQKRPGRVIRDKRLRYAHC 267  
 DB 419 GAIRINTQEVLDVLRPMVLADKAKQVQWTVNYLDLLEL-----GLVITGTLPEVNTIG 474  
 QY 268 HHNNYVSGSTITFIINLHRSRKKIKLAGTLRDLKHQYLQYPYQEGLSKSKSVQLNGQP 327  
 DB 475 FENTNLKNQILGVMGVDVSLEDIK-----RLTPRFTLCPNGY----YFAIDPN 523  
 QY 328 LVMVDDGTLPKLPRLRAGRTLVP-----PVTMGFFVFNVNVALA 369  
 DB 524 LLH-----PNLQPKPIGVG----IPTNLKRRPNVQNPQSKBPVTLDFLDABLENDIK 573  
 QY 370 CRYR 373  
 DB 574 VEIR 577  
 RESULT 13  
 COXI\_SACDO  
 ID COXI\_SACDO STANDARD; PRT; 534 AA.  
 AC P98001;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).  
 GN COXI.  
 OS Saccharomyces douglasii (Yeast).  
 OG Mitochondrion.  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=46617;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SD12;  
 RX MEDLINE=93185919; PubMed=8383070;  
 RA Tian G.L., Michel F., Macadre C., Lazowska J.;  
 RT "Sequence of the mitochondrial gene encoding subunit I of cytochrome  
 RL oxidase in Saccharomyces douglasii.";  
 RL Gene 124:153-163(1993).  
 CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE



FT	METAL	383	383	IRON (HEME A) (PROBABLE).
FT	LIPID	324	324	MYRISTATE.
SO	SEQUENCE	557 AA:	61494 MW:	D4D9618704E5DF6C CRC64;

	Query Match	4.5%;	Score 87.5;	DB 1;	Length 557;
	Best Local Similarity	22.0%;	Pred. No. 11;		
	Matches	46;	Conservative	33;	Mismatches 69; Indels 61; Gaps 13;
QY	74	IRIYSRASLYGPNIGRPRKNVIALDGFMKVAGSTVDATW-QHCYIDGRVVKVMDFLKT	132		
Db	264	ISAYSNKSVFG-----YIGVYAMMSI--GILGFIWSHHMYTVG-----LD-VDT	306		
QY	133	RLDLTLDQIRKIQVNTYTPGKKINLEGVVTVTSAGGTNNLSDS--YAAGFLMLNTLIG-	189		
Db	307	RAYETAATLIIAVPTGIKIFS-----WL---ATCYGGSIRLTPSMLFALGFVFMFTIGG	357		
QY	190	---MLANOGIDVYIRUSFF-----DHGYNHLVDQ---NFNPL---P	221		
Db	358	LSGVVLNANLIDIAFHDTYYYVAHFHYVLSMGAVFMFSGWHVVPKILGLNYNNVLSKA	417		
QY	222	DYWLSSLYKRLIGPKVLAVHVAGLQRPK	250		
Db	418	OFWLLFTGVNL---TFEPOHFLGLOGMPP	443		

```

RESULT 15
MUTL_BACSU STANDARD; PRP; 627 AA.
ID P49850;
AC DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA mismatch repair protein mutL.
GN MUTL.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
[ ]
RN SEQUENCE FROM N.A.
RP STRAIN=168;
RC MEDLINE=96349107; PubMed=8760914;
RX Ginetti F., Perego M., Albertini A.M., Galizzi A.;
RT "Bacillus subtilis mutS mutL operon: identification, nucleotide
sequence and mutagenesis.";
RL Microbiology 142:2021-2029(1996).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES IN
DNA. IT IS REQUIRED FOR DAM-DEPENDENT METHYL-DIRECTED DNA MISMATCH
REPAIR. MAY ACT AS A "MOLECULAR MATCHMAKER", A PROTEIN THAT
PROMOTES THE FORMATION OF A STABLE COMPLEX BETWEEN TWO OR MORE
DNA-BINDING PROTEINS IN AN ATP-DEPENDENT MANNER WITHOUT ITSELF
BEING PART OF A FINAL EFFECTOR COMPLEX (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTL/HEXB FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U27343; AAB19236.1; -.
DR EMBL; Z99112; CAB13578.1; -.
DR HSSP; P23367; LBKN.
DR Subtilist; BG11402; mutL.
DR InterPro; IPR002099; DNA_mis_repair.
DR InterPro; IPR003594; HATPase_c.
DR InterPro; IPR004359; HIS_KIN_sig.
DR Pfam; PF01119; DNA_mis_repair; 1.
DR Pfam; PF02518; HATPase_c; 1.
DR ProSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.

```



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:30:49 ; Search time 123.94 Seconds  
(without alignments)  
520.632 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_162\_534  
Perfect score: 1956  
Sequence: 1 KGCKTAQHPDVMLEQREKA.....PVTMGFFVKNYNALACRYR 373

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPREMBL\_19.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mmc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1956	100.0	534	4 Q9HB38	Q9hb38 homo sapien
2	1917	98.0	592	4 Q9HB37	Q9hb37 homo sapien
3	1743	89.1	480	4 Q9HB39	Q9hb39 homo sapien
4	810.5	41.4	536	11 Q9QZf8	Q9qzf8 rattus norv
5	787.5	40.3	543	4 Q9Y251	Q9y251 homo sapien
6	787.5	40.3	545	4 Q9UL39	Q9ul39 homo sapien
7	778	39.8	545	6 Q9MYV0	Q9myv0 bos taurus
8	736.5	37.7	523	13 Q90YK5	Q90yk5 gallus gall
9	301	15.4	521	10 Q9SDA1	Q9sda1 arabisopsis
10	301	15.4	543	10 Q9FF10	Q9ff10 arabisopsis
11	258.5	13.2	516	10 Q9FLK8	Q9flk8 arabisopsis
12	236	12.1	536	10 Q9FZP1	Q9fzp1 arabisopsis
13	215.5	11.0	527	10 Q9LRC8	Q9lrc8 scutellaria
14	155	7.9	190	10 O82604	O82604 arabisopsis
15	155	7.9	935	5 Q9VE79	Q9ve79 drosophila
16	110	5.6	174	10 Q9ATW5	Q9atw5 zea mays (m

ALIGNMENTS

RESULT 1

Q9HB38	102.5	5.2	1260	5	015639	015639 dictyosteli
ID Q9HB38	98	5.0	491	2	008457	008457 clostridium
AC Q9HB38;	96.5	4.9	1429	2	09LAP7	09lap7 alteromonas
DT 01-MAR-2001 (TREMBLrel. 16, Created)	95.5	4.9	551	10	Q38779	Q38779 avena sativ
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)	95.5	4.9	663	8	Q9MJ69	Q9mj69 physarum po
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)	95	4.9	512	8	Q99806	Q99806 ixodes hexa
DE HEPARANASE-LIKE PROTEIN HPA2B.	94.5	4.8	460	16	Q98QL8	Q98ql8 mycoplasma
OS Homo sapiens (Human).	94	4.8	855	15	Q90DZ7	Q90dz7 human immu
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	94	4.8	1234	5	Q9YLH7	Q9ylh7 dictyosteli
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	93	4.8	320	17	Q97UM5	Q97um5 sulfolobus
OX NCBI_TaxID=9606;	93	4.8	441	10	Q94L11	Q94l11 arabisopsis
RN [1]	93	4.8	515	10	Q49258	Q49258 avena sativ
RP MEDLINE=20483645; PubMed=11027606;	92.5	4.7	620	10	Q9FU14	Q9fu14 oryza sativ
RX McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,	92.5	4.7	876	16	Q99TH2	Q99th2 staphylococ
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;	92.5	4.7	1845	12	Q9WJZ8	Q9wjz8 human cytom
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian	92.5	4.7	585	10	Q9FJ99	Q9fj99 arabisopsis
RT Heparanase Family Member.;"	90.5	4.6	831	11	Q9ESY7	Q9esy7 mus musculu
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).	90.5	4.6	831	11	Q9LWG4	Q9lwg4 mus musculu
DR EMBL; AF282886; AAG23422.1; -	90.5	4.6	876	2	Q9RMM6	Q9rm6 staphylococ
SQ SEQUENCE 534 AA; 60063 MW; C3DE5E900CB338C4 CRC64;	90	4.6	471	8	Q9MH11	Q9mhl1 grevillea r

Query Match	100.0%	Score	1956;	DB	4;	Length	534;
Best Local Similarity	100.0%	Pred. No.	3.7e-162;				
Matches	373;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	KGCKTAQHPDVMLEQREKAQMHLVLLKQFSNTYSNLILTEPNRYTMHGRAVNSQL	60				
Db	162	KGCKTAQHPDVMLEQREKAQMHLVLLKQFSNTYSNLILTEPNRYTMHGRAVNSQL	221				
Qy	61	KGKDIQLKSLQPIRIYRSASLYGNIGPRKNVIALLDGFMKVGSTVDATVTHQHCYD	120				
Db	222	KGKDIQLKSLQPIRIYRSASLYGNIGPRKNVIALLDGFMKVGSTVDATVTHQHCYD	281				
Qy	121	GRVVKVMDFLKRLDLSQIRKTKQVNTYTPGKKIWLGVVTSAGGTNNLSDSVA	180				
Db	282	GRVVKVMDFLKRLDLSQIRKTKQVNTYTPGKKIWLGVVTSAGGTNNLSDSVA	341				

QY 181 GFLWNTLGLMANGIDVIRHSFFDHGYNHLVDQNFNPLDPYWLKRLIGPKVLAV 240  
Db 342 GFLWNTLGLMANGIDVIRHSFFDHGYNHLVDQNFNPLDPYWLKRLIGPKVLAV 401  
QY 241 HVAGLQKRPGRVIRDKLRIYAHCTNHHNNHNVVRSITLFIINLHRSRKKIKLAGTLRD 300  
Db 402 HVAGLQKRPGRVIRDKLRIYAHCTNHHNNHNVVRSITLFIINLHRSRKKIKLAGTLRD 461  
QY 301 KLHVQYLLQPYGQGLKSKSVQNLGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFF 360  
Db 462 KLHVQYLLQPYGQGLKSKSVQNLGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFF 521  
QY 361 VVKVNNALACRYR 373  
Db 522 VVKVNNALACRYR 534  
RESULT 2  
ID Q9HB37 PRELIMINARY; PRT; 592 AA.  
AC Q9HB37;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE HEPARANASE-LIKE PROTEIN HPA2C.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20483645; PubMed=11027606;  
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,  
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;  
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian  
RT Heparanase Family Member."  
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).  
DR EMBL: AF282887; AAG23423.1; -.  
SQ SEQUENCE 592 AA; 66580 MW; 95C384AD9A74258E CRC64;

Query Match 98.0%; Score 1917; DB 4; Length 592;  
Best Local Similarity 86.5%; Pred. No. 1.1e-158;  
Matches 373; Conservative 0; Mismatches 0; Indels 58; Gaps 1;  
QY 1 KGCKIAQHPDVMLEQREKAAQHLLVLLKEQFSNTYSNLIIT----- 42  
Db 162 KGCKIAQHPDVMLEQREKAAQHLLVLLKEQFSNTYSNLIITARSIDKLYNFADCSGLHL 221  
QY 43 -----EPNNYRTMHGRAVNGSQLGK 62  
Db 222 IFALNLRNPNNSWSSALSLLKYSASKYINISHELGNENPNYRTMHGRAVNGSQLGK 281  
QY 63 DYIQLKSLQPIRIYRASLYGNIGRPRKNVIALDGMKVGSTVDVAVTWHQCHYIDGR 122  
Db 282 DYIQLKSLQPIRIYRASLYGNIGRPRKNVIALDGMKVGSTVDVAVTWHQCHYIDGR 341  
QY 123 VVKVNDFLKRLDITLSDQIRKTKQKVNTYTPGKKIWLGVVTTISAGGTNNLSDSYAAGF 182  
Db 342 VVKVNDFLKRLDITLSDQIRKTKQKVNTYTPGKKIWLGVVTTISAGGTNNLSDSYAAGF 401  
QY 183 LWLNTLGLMANGIDVIRHSFFDHGYNHLVDQNFNPLDPYWLKRLIGPKVLAVHV 242  
Db 402 LWLNTLGLMANGIDVIRHSFFDHGYNHLVDQNFNPLDPYWLKRLIGPKVLAVHV 461  
QY 243 AGLQKRPGRVIRDKLRIYAHCTNHHNNHNVVRSITLFIINLHRSRKKIKLAGTLRDKL 302  
Db 462 AGLQKRPGRVIRDKLRIYAHCTNHHNNHNVVRSITLFIINLHRSRKKIKLAGTLRDKL 521  
QY 303 VHQLLPQYGGEGKSKSVQNLGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFFV 362  
Db 522 VHQLLPQYGGEGKSKSVQNLGQPLVMVDDGTLPELKPRLRAGRTLVIPTVMGFFV 581

QY 363 KVNVALACRYR 373  
Db 582 KVNVALACRYR 592  
RESULT 3  
ID Q9HB39 PRELIMINARY; PRT; 480 AA.  
AC Q9HB39;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE HEPARANASE-LIKE PROTEIN HPA2A.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20483645; PubMed=11027606;  
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,  
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;  
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian  
RT Heparanase Family Member."  
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).  
DR EMBL: AF282885; AAG23421.1; -.  
SQ SEQUENCE 480 AA; 53900 MW; F75F89F67AC1FF83 CRC64;  
Query Match 89.1%; Score 1743; DB 4; Length 480;  
Best Local Similarity 100.0%; Pred. No. 1.2e-143;  
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 43 EPNNYRTMHGRAVNGSQLGKDYIQLKSLQPIRIYRASLYGNIGRPRKNVIALDGM 102  
Db 150 EPNNYRTMHGRAVNGSQLGKDYIQLKSLQPIRIYRASLYGNIGRPRKNVIALDGM 209  
QY 103 KVAGSTVDVAVTWHQCHYIDGRVVKVNDFLKRLDITLSDQIRKTKQKVNTYTPGKKIWL 162  
Db 210 KVAGSTVDVAVTWHQCHYIDGRVVKVNDFLKRLDITLSDQIRKTKQKVNTYTPGKKIWL 269  
QY 163 VVTSAGTNNLSDSYAAGFLWNTLGLMANGIDVIRHSFFDHGYNHLVDQNFNPLD 222  
Db 270 VVTSAGTNNLSDSYAAGFLWNTLGLMANGIDVIRHSFFDHGYNHLVDQNFNPLD 329  
QY 223 YWLSLLYKRLIGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNHHNNHNVVRSITLFI 282  
Db 330 YWLSLLYKRLIGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNHHNNHNVVRSITLFI 389  
QY 283 INLHRSRKKIKLAGTLRDKLHVQYLLQPYGQGLKSKSVQNLGQPLVMVDDGTLPELKP 342  
Db 390 INLHRSRKKIKLAGTLRDKLHVQYLLQPYGQGLKSKSVQNLGQPLVMVDDGTLPELKP 449  
QY 343 PLRAGRTLVIPTVMGFFVVKVNNALACRYR 373  
Db 450 PLRAGRTLVIPTVMGFFVVKVNNALACRYR 480  
RESULT 4  
ID Q90ZF8 PRELIMINARY; PRT; 536 AA.  
AC Q90ZF8;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE HEPARANASE.  
OS HEP.  
GN Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]



[illegible]

AC O9SDAL: 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE HYPOTHEICAL 57.8 KDA PROTEIN.  
GN F13G24.30.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Van Der Schueren J., Chuang Y.J., Voet M., Robben J.,  
RA Volckaert G., Bancroft I., Meves H.W., Lemcke K., Mayer K.F.X.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL133421; CAB62595.1; -.  
DR InterPro; IPR001254; Trypsin.  
DR PROSITE; PS00135; TRYPsin\_SER; UNKNOWN\_1.  
KW Hypothetical protein.  
SQ SEQUENCE 521 AA; 57831 MW; 07D8664A4B305CC2 CRC64;  
  
Query Match 15.4%; Score 301; DB 10; Length 521;  
Best Local Similarity 28.2%; Pred. No. 9.8e-18;  
Matches 105; Conservative 50; Mismatches 127; Indels 90; Gaps 15;  
  
QY 52 GRAVNGSQLGKDYIQLKSLLOPIRIYSRASLYGPNIGRPRKNVIALLDGF-----M 102  
Db 186 GASVSAELYGKDLIVLKDVIN--KVYKNSWLHKPILVAP-----GGFYEQQWTKLL 235  
QY 103 KVAG-STVDVAVTQHCYIDGR-----VVKVMDFLKTRLLDLSQIRKIQKVNTYTPG 155  
Db 236 EISGPSVDVWT-HHIYNLGSNDPALVKKIMD---PSYLSQVSKTFKDVNQTIQEHGP- 290  
QY 156 KKIWLEGVVTTSAG---GTNNLSDSYAAGFLWNLTLGMLANOGIDVIRHSHFFDHGYNH 211  
Db 291 ---WASPWVGSGGAYNSGGRHVSTFIDSFYLDQLGMSARHNTKVYCRQLVGGFYGL 347  
QY 212 LVDQNFNPLPDWLSLLKRLGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNHHNH 271  
Db 348 LEKGFVFPNPDYISALLWHLRMKGVLAVQTDG-----PPQLRVYAHCSKG--- 393  
QY 272 NVYRGSIITLFIINLH-----RSRKKIKLAGTLRDLKLV----- 303  
Db 394 ---RAGVTLLINLSNQSDFTVSVSNGINVLNAESRKKKSLDITLKRPFWSIGSKASDG 450  
QY 304 ----HQYLLQPYGQEG-LKSKSVOLNGOPLVMVDDGTLPKLPRLRAGRTLVIPVPTWG 358  
Db 451 YLNREYHLP--ENGVLRSKTMVLNGSKLKTATGDIPLSLEPVLRSVNSPLNVLPLSMS 508  
QY 359 FFWKVNVALAC 370  
Db 509 FIVLPNFDASAC 520  
  
RESULT 10  
Q9FF10 ID Q9FF10 PRELIMINARY; PRT; 543 AA.  
AC Q9FF10;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE SIMILARITY TO HEPARANASE.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=97471969; PubMed=9330910;  
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,  
RA Miyajima N., Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence  
RT features of the 1.6 Mb regions covered by twenty physically assigned  
RT pl clones";  
RL DNA Res. 4:215-230(1997).  
DR EMBL; AB005249; BAB09947.1; -.  
DR InterPro; IPR001254; Trypsin.  
DR PROSITE; PS00135; TRYPsin\_SER; UNKNOWN\_1.  
SQ SEQUENCE 543 AA; 60250 MW; 0FA2248948282FF6 CRC64;  
  
Query Match 15.4%; Score 301; DB 10; Length 543;  
Best Local Similarity 28.2%; Pred. No. 1e-17;  
Matches 105; Conservative 50; Mismatches 127; Indels 90; Gaps 15;  
  
QY 52 GRAVNGSQLGKDYIQLKSLLOPIRIYSRASLYGPNIGRPRKNVIALLDGF-----M 102  
Db 208 GASVSAELYGKDLIVLKDVIN--KVYKNSWLHKPILVAP-----GGFYEQQWTKLL 257  
QY 103 KVAG-STVDVAVTQHCYIDGR-----VVKVMDFLKTRLLDLSQIRKIQKVNTYTPG 155  
Db 258 EISGPSVDVWT-HHIYNLGSNDPALVKKIMD---PSYLSQVSKTFKDVNQTIQEHGP- 312  
QY 156 KKIWLEGVVTTSAG---GTNNLSDSYAAGFLWNLTLGMLANOGIDVIRHSHFFDHGYNH 211  
Db 313 ---WASPWVGSGGAYNSGGRHVSTFIDSFYLDQLGMSARHNTKVYCRQLVGGFYGL 369  
QY 212 LVDQNFNPLPDWLSLLKRLGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNHHNH 271  
Db 370 LEKGFVFPNPDYISALLWHLRMKGVLAVQTDG-----PPQLRVYAHCSKG--- 415  
QY 272 NVYRGSIITLFIINLH-----RSRKKIKLAGTLRDLKLV----- 303  
Db 416 ---RAGVTLLINLSNQSDFTVSVSNGINVLNAESRKKKSLDITLKRPFWSIGSKASDG 472  
QY 304 ----HQYLLQPYGQEG-LKSKSVOLNGOPLVMVDDGTLPKLPRLRAGRTLVIPVPTWG 358  
Db 473 YLNREYHLP--ENGVLRSKTMVLNGSKLKTATGDIPLSLEPVLRSVNSPLNVLPLSMS 530  
QY 359 FFWKVNVALAC 370  
Db 531 FIVLPNFDASAC 542  
  
RESULT 11  
Q9FLK8 ID Q9FLK8 PRELIMINARY; PRT; 516 AA.  
AC Q9FLK8;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE SIMILARITY TO HEPARANASE.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=98290546; PubMed=9628582;  
RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.  
RT Sequence features of the regions of 1,456,315 bp covered by nineteen  
RT physically assigned pl and TAC clones";  
RL DNA Res. 5:41-54(1998).  
DR EMBL; AB010073; BAB08480.1; -.



```
RESULT 14
082604 ID O82604 PRELIMINARY; PRT; 190 AA.
AC O82604;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE T2L5.6 PROTEIN.
GN T2L5.6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Geisel C., Smith A., Le T.;
RT "The sequence of A. thaliana T2L5.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA WashU;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Waterston R.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF096371; AAC62794.1; -.
SQ SEQUENCE 190 AA; 21212 MW; B5E356E526007A0 CRC64;

Query Match 7.9%; Score 155; DB 10; Length 190;
Best Local Similarity 24.9%; Pred. No. 1.2e-05;
Matches 49; Conservative 32; Mismatches 70; Indels 46; Gaps 6;

QY 201 RHSEFDHGHVHLVDQNFNLPDVLWLSLLYKRLGPKVLAVHVGALQKPPRGRVIRDKLR 260
DB 12 RSLGIGNTGGLTNTFTNPDIYSALWQLMGRKALFTTFSGTK-----KIR 60

QY 261 IYAHCTNNHNNHNVGRSGITLFIINLHRSR---KKIKLAGTLRDKLVHQVLLQPY----- 311
DB 61 SYTHCARQSK-----GITVLLMLNDNTTIVVAKVELNNSF--SLRTRKHKSKYKRASSQ 112

QY 312 ---GQEG-----LKSQVQLNGQPLVWDDGTLPKLPRLRAGRTLVIP 353
DB 113 LFGGPGNGVIOREYHYHTAKDGNLHQSOTMLNGNALQVNSMGDLPPIEPIHINSTEPTIA 172

QY 354 PVTMGFFVKNVNNALAC 370
DB 173 PYSIVEVHRNVVVPAC 189

RESULT 15
Q9VE79 ID Q9VE79 PRELIMINARY; PRT; 935 AA.
AC Q9VE79;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE CG14309 PROTEIN.
GN CG14309.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003721; AAF55548.1; -.
DR FlyBase; FBgn0038611; CG14309.
SQ SEQUENCE 935 AA; 106883 MW; 10B46B721CE5DB3F CRC64;
```

```
Query Match 7.9%; Score 155; DB 5; Length 935;
Best Local Similarity 22.8%; Pred. No. 0.00013;
Matches 84; Conservative 54; Mismatches 130; Indels 104; Gaps 21;
```

```
QY 44 PNYRTMHG-RAYNGS-----OLG-----KDYIQ-----LKSLLQPIRIY-SR 79
DB 85 PDDWDSMHTLKILNTSYMVGITDCINQLGTFDGTSGRAKDYVQELRTLKLWDTFKPYDD 144

QY 80 ASLYGPNIGRPNKRNIALLDGFMKVGAGTVDAVWQHCHYID-----GRVVKMDFL 130
DB 145 WRLMGADIS-----AGSSADET---KRYVDMSKDLNTAFGQTQANNLP 185

QY 131 KTRILDTLSDQIRKIOKVNTYTPGKKIWLEGVYVTTSGGTNNLSDSYAGFLWLNTLGM 190
DB 186 KSSLGSYLYDSDPALRTLQOORVP---LWL--TLPEERSQRLVGDETTDALRWAQTMGD 240

QY 191 LANQGDVWVIRHSFFDHGYNHLVD---QNFENPLPDVWLSLLYKRLGPKVLAVHVGALQR 247
DB 241 AAASGDFVIEFRM-----NLVDFERPNS-----LYVTALFKYMGSRVFPA----- 282

QY 248 KPRGVRIRDKLRIYAHCTNNHNNHNVGRSGITLFIINLHRSRKKI--KLAGTLRDKLVHQ 305
DB 283 --RPLNAFAPSNKLYTHCA-----NAVSGGLAFWNVTEEQPTITIVKSTSSLSSSBIWQ 335

QY 306 YLLQPTGQGLKSKSVQLNGQPLVWDDGTL-PELKP-----RPLRAGRTLVIPIVPMGFF 360
DB 336 YVLTGHDQR-----VOLNNVRLHL--NTTLRLPLIKPIDTKPLQ-----LITFSMAVSFW 383
```

.Qy 361 VVKNNVALACRY 372  
|: :|| |:  
Db 384 VLPDVNLEHCQF 395

Search completed: July 30, 2002, 08:30:51  
Job time: 1003 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 08:15:07 ; Search time 53.59 Seconds  
(without alignments)  
170.008 Million cell updates/sec

Title: US-09-836-461-2\_COPY\_162\_534

Perfect score: 1956

Sequence: 1 KGCKIAQHPDVMLQLEKA.....PVTMGFFVVKVYNALACRYR 373

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents\_AA:\*
- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
  - 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
  - 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*
  - 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	805	41.2	380	4	US-09-181-336-19
2	791.5	40.5	543	2	US-08-922-1708-10
3	791.5	40.5	543	4	US-09-071-7398-2
4	791.5	40.5	543	4	US-09-260-038B-2
5	787.5	40.3	543	4	US-09-181-336-13
6	779	39.8	380	4	US-09-181-336-17
7	744.5	38.1	532	4	US-09-181-336-15
8	88	4.5	1106	1	US-08-435-675B-5
9	88	4.5	1106	1	US-08-336-257A-8
10	86	4.4	1103	1	US-08-455-543A-53
11	86	4.4	1103	2	US-08-223-305C-53
12	84.5	4.3	434	2	US-08-989-925-1
13	84	4.3	538	4	US-09-175-928-4
14	82.5	4.2	921	1	US-07-872-644-39
15	82.5	4.2	921	1	US-08-297-494-39
16	82.5	4.2	921	1	US-08-297-510-39
17	82.5	4.2	921	1	US-08-479-532-39
18	82.5	4.2	921	1	US-08-455-526-39
19	82.5	4.2	921	1	US-08-455-525-39
20	82.5	4.2	921	3	US-09-139-491-39
21	82.5	4.2	921	5	PCT-US92-03222-39
22	82.5	4.2	942	1	US-07-872-644-43
23	82.5	4.2	942	1	US-08-297-494-43
24	82.5	4.2	942	1	US-08-297-510-43
25	82.5	4.2	942	1	US-08-479-532-43
26	82.5	4.2	942	1	US-08-455-526-43
27	82.5	4.2	942	1	US-08-455-525-43

28	82.5	4.2	942	3	US-09-139-491-43	Sequence 43, Appl
29	82.5	4.2	942	5	PCT-US92-03222-43	Sequence 43, Appl
30	82	4.2	591	3	US-09-082-737-2	Sequence 2, Appl
31	81	4.1	685	2	US-08-878-989-1	Sequence 1, Appl
32	81	4.1	685	3	US-09-136-282-2	Sequence 2, Appl
33	81	4.1	685	4	US-09-272-796-1	Sequence 1, Appl
34	81	4.1	685	4	US-09-505-744-2	Sequence 2, Appl
35	81	4.1	1788	2	US-08-962-284-2	Sequence 2, Appl
36	80.5	4.1	1086	6	5386025-8	Patent No. 5386025
37	80	4.1	313	1	US-08-302-449-2	Sequence 2, Appl
38	80	4.1	313	5	PCT-US94-07430-2	Sequence 2, Appl
39	80	4.1	1093	5	PCT-US93-03077-1	Sequence 1, Appl
40	79	4.0	934	1	US-08-215-805A-80	Sequence 80, Appl
41	79	4.0	1541	4	US-08-296-791-3	Sequence 3, Appl
42	79	4.0	1541	5	PCT-US95-10561A-3	Sequence 3, Appl
43	78.5	4.0	313	1	US-08-302-449-4	Sequence 4, Appl
44	78.5	4.0	313	5	PCT-US94-07430-4	Sequence 4, Appl
45	78.5	4.0	687	5	PCT-US91-09784-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-181-336-19  
; Sequence 19, Application US/09181336A  
; Patent No. 6242238  
; GENERAL INFORMATION:  
; APPLICANT: FREEMAN, Craig Geoffrey  
; APPLICANT: PARISH, Richard  
; APPLICANT: HUMORE, Brenton James  
; APPLICANT: HULETT, Mark Darren  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN  
; TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR  
; FILE REFERENCE: 032505-004  
; CURRENT APPLICATION NUMBER: US/09/181.336A  
; CURRENT FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: AU PP0062  
; EARLIER FILING DATE: 1997-10-28  
; EARLIER APPLICATION NUMBER: AU PP0812  
; EARLIER FILING DATE: 1997-12-09  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Rattus sp.  
US-09-181-336-19

Query Match 41.2%; Score 805; DB 4; Length 380;

Best Local Similarity 47.3%; Pred No. 5.1e-80;

Matches 155; Conservative 57; Mismatches 106; Indels 10; Gaps 2;

QY	43	EPNNYRTMHGRAVNSQLGKDYIOLKSLIQTIRIYSRASLYGNIGRPRKNVIALLDGFM	102
Db	62	EPNSFWKKAHISIDGLQGLGDFVHLKLLQK-SAFQNAKLYGPDIGQPRGKTVKLLRSFL	120
QY	103	KVAGSTVDATVWQHGYIDGRVYKVMDFLKLRLDLSDOIRKIQKVVNTYTPGKKIWLEG	162
Db	121	KAGGVEIDSLTWHHYLYLNGRVATKEDFLSSDVLDTFILSVQKILKVTBMTGKKVWLE	180
QY	163	VYTTSGAGTNNLSDSYAAGFLMLNTLGMLANOGIDWIRHSFFDGHYNHLVDQNFNPLPD	222
Db	181	TSSAYGGGAPLSDITFAGFWMLDKLGISAQLGIEVVMHQVFFGAGNYHLVDENEPEPLD	240
QY	223	YWLSTLYKRLIGPKVLAVHAGLQRPGRVIRDKRLIYAHCTNHHNNHYVRSITLFI	282
Db	241	YWLSLLFKKLVGPKVLMRSVKGPD-----RSKRLVYLHCTNVYHPRYREGDITLVY	291
QY	283	INLHRSRKIKLAGLIRDKLVHQYLLQPYQGGGLSKSVQLNGQPLVWVDDGTLPKPR	342
Db	292	LNLHNVTKLKPPPMFSRPVDKYLKLPFGSDGLLSKSVQLNGQTLKMVDEQTLPALTEK	351



Db 317 DIFISSVQKVFQVVESTRPGKKNVWLGTSAYGGGAPLSDTFAAGFWMWDLKGLSARMG 376  
QY 196 IDVYIRHSFFDHGYNHLVDQNFNPLPDYWLSSLLYKRLGPKVLAVHVAGLQKRPGRVI 255  
Db 377 IEVVMQVFFGAGNHLVDENFDPLPDYWLSSLLPKKLVGTAKVLMASVQSGRR----- 429  
QY 256 RDKLRIYAHCTNHHNNHNVVRSITLFIINLHRSRKKIKLAGTLRDKLVHQLQPYGQEG 315  
Db 430 --KLRYLHCTNTPRYKEGDLTLAIINLHNVTKYLRPLPFFSNKQVDKYLRLPLGPHG 487  
QY 316 LKSKSVOLNGLOPLVMVDDGTLPKLPRLAGRTLVIPVMTGFFVKNVNALAC 370  
Db 488 LLSKSVQLNGTLKMWDDQTLPLMEKPLRGSSGLGPAFSYFFVIRNAKVAAC 542

RESULT 4  
US-09-260-038B-2  
; Sequence 2, Application US/09260038B  
; Patent No. 6348344  
; GENERAL INFORMATION:  
; APPLICANT: Maty Ayal-Hershkovitz et al.  
; TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS FOR  
; EXPRESSING RECOMBINANT HEPARANASE  
; AND METHODS OF PURIFYING SAME  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina  
; STREET: 2001 Jefferson Davis Highway, Suite 207  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: United States of America  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk  
; COMPUTER: Twinhead\* Slimnote-890TX  
; OPERATING SYSTEM: MS DOS version 6.2,  
; Windows version 3.11  
; SOFTWARE: Word for Windows version 2.0 converted to  
; an ASCII file  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/260,038B  
; FILING DATE: 02-Mar-1999  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/071,618  
; FILING DATE: May 1, 1998  
; APPLICATION NUMBER: 09/071,739  
; FILING DATE: May 1, 1998  
; APPLICATION NUMBER: 08/922,180  
; FILING DATE: September 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Friedman, Mark M.  
; REGISTRATION NUMBER: 33,883  
; REFERENCE/DOCKET NUMBER: 910/16  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 972-3-562553  
; TELEFAX: 972-3-562554  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 543  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-260-038B-2

Query Match 40.5%; Score 791.5; DB 4; Length 543;  
Best Local Similarity 44.2%; Pred. No. 2.7e-78;  
Matches 157; Conservative 58; Mismatches 115; Indels 25; Gaps 3;

QY 31 QFSNTYSNLIL-----TEPNYRTMHGRAVNGSQLGKDYIQLKSLQPIR 75  
Db 198 QWNSSNAQLLLDYCSSKGYNISWELGNENPSFLKKADIFINGSQLGEDYIQLHKLRLK-S 256  
QY 76 IYSRASLYGPNIGRPNKKNVIALLDGFMKVAGSTVDATVWQHGYIDGRVVYKVMDFLKTLL 135  
Db 257 TEKNAKLYGPDVGQPRRTAKMKSFLKAGGEVIDSVTWHHYLLNGRTATREDFLNPDVL 316  
QY 136 DTLSDQIRKIQKVVNTYTPGKKIWLGVVTTTSGAGTNNLSDSYAAGFLWNLTLGLMANOG 195  
Db 317 DIFISSVQKVFQVVESTRPGKKNVWLGTSAYGGGAPLSDTFAAGFWMWDLKGLSARMG 376  
QY 196 IDVYIRHSFFDHGYNHLVDQNFNPLPDYWLSSLLYKRLGPKVLAVHVAGLQKRPGRVI 255  
Db 377 IEVVMQVFFGAGNHLVDENFDPLPDYWLSSLLPKKLVGTAKVLMASVQSGRR----- 429  
QY 256 RDKLRIYAHCTNHHNNHNVVRSITLFIINLHRSRKKIKLAGTLRDKLVHQLQPYGQEG 315  
Db 430 --KLRYLHCTNTPRYKEGDLTLAIINLHNVTKYLRPLPFFSNKQVDKYLRLPLGPHG 487  
QY 316 LKSKSVOLNGLOPLVMVDDGTLPKLPRLAGRTLVIPVMTGFFVKNVNALAC 370  
Db 488 LLSKSVQLNGTLKMWDDQTLPLMEKPLRGSSGLGPAFSYFFVIRNAKVAAC 542

RESULT 5  
US-09-181-336-13  
; Sequence 13, Application US/09181336A  
; Patent No. 6242238  
; GENERAL INFORMATION:  
; APPLICANT: FREEMAN, Craig Geoffrey  
; APPLICANT: PARISH, Richard  
; APPLICANT: HAMDORF, Brenton James  
; APPLICANT: HULETT, Mark Darren  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN  
; TITLE OF INVENTION: ENDOGLUCURONIDASE AND USES THEREFOR  
; FILE REFERENCE: 032503-004  
; CURRENT APPLICATION NUMBER: US/09/181,336A  
; CURRENT FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: AU PP0062  
; EARLIER FILING DATE: 1997-10-28  
; EARLIER APPLICATION NUMBER: AU PP0812  
; EARLIER FILING DATE: 1997-12-09  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 543  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-181-336-13

Query Match 40.3%; Score 787.5; DB 4; Length 543;  
Best Local Similarity 43.9%; Pred. No. 7.6e-78;  
Matches 156; Conservative 59; Mismatches 115; Indels 25; Gaps 3;

QY 31 QFSNTYSNLIL-----TEPNYRTMHGRAVNGSQLGKDYIQLKSLQPIR 75  
Db 198 QWNSSNAQLLLDYCSSKGYNISWELGNENPSFLKKADIFINGSQLGEDYIQLHKLRLK-S 256  
QY 76 IYSRASLYGPNIGRPNKKNVIALLDGFMKVAGSTVDATVWQHGYIDGRVVYKVMDFLKTLL 135  
Db 257 TEKNAKLYGPDVGQPRRTAKMKSFLKAGGEVIDSVTWHHYLLNGRTATREDFLNPDVL 316  
QY 136 DTLSDQIRKIQKVVNTYTPGKKIWLGVVTTTSGAGTNNLSDSYAAGFLWNLTLGLMANOG 195  
Db 317 DIFISSVQKVFQVVESTRPGKKNVWLGTSAYGGGAPLSDTFAAGFWMWDLKGLSARMG 376  
QY 196 IDVYIRHSFFDHGYNHLVDQNFNPLPDYWLSSLLYKRLGPKVLAVHVAGLQKRPGRVI 255  
Db 377 IEVVMQVFFGAGNHLVDENFDPLPDYWLSSLLPKKLVGTAKVLMASVQSGRR----- 429  
QY 256 RDKLRIYAHCTNHHNNHNVVRSITLFIINLHRSRKKIKLAGTLRDKLVHQLQPYGQEG 315

Db 430 --KLVYLHCTNTDPRYKEGDLTVYALNHNVTYRLPYFSPKQVDKYLRLPLGPHG 487  
QY 316 LKSKSVQLNGQPLVMVDGTLPELKPRLRAGRTLVIPTVMGFVVKVNNALAC 370  
Db 488 LLSKSVQLNGTLTKMVDQTLPLMEKPLRPGSSGLGPAFSYFVIRNAKYAAC 542

## RESULT 6

US-09-181-336-17  
; Sequence 17, Application US/09181336A  
; Patent No. 6242238  
; GENERAL INFORMATION:  
; APPLICANT: FREEMAN, Craig Geoffrey  
; APPLICANT: PARISH, Richard  
; APPLICANT: HAMDORF, Brenton James  
; APPLICANT: HULETT, Mark Darren  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN  
; FILE REFERENCE: 032505-004  
; CURRENT APPLICATION NUMBER: US/09/181.336A  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER FILING DATE: 1997-10-28  
; EARLIER FILING DATE: 1997-12-09  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-181-336-17

Query Match 39.8%; Score 779; DB 4; Length 380;

Best Local Similarity 46.3%; Pred. No. 3.7e-77;  
Matches 152; Conservative 56; Mismatches 110; Indels 10; Gaps 2;

QY 43 EPNVYRTMHGRAVNSOLGKDYIQLKSLLOPIRIYSRASLYGNIGRPNKVNIALLDGFM 102  
Db 62 EPNSEWKAHILIDGLQGEDEVEHLKLLQR-SAFONAKLYGPDIGQPRGKTVKLLRSFL 120  
QY 103 KVAGSTDAVTWQHGYIDGRVVKVMDFLKTRLLDLSQIRKIQKVNTYTPGKKIWLEG 162  
Db 121 KAGGEVIDSLTHWHYYLNGRIATKEDFLSSDVLDTFILSVQKILKVTKEITPGKKVWLE 180  
QY 163 VYTTAGGTTNNLSYAGFLWNTLGMANOGIDVWIRHSFFDHGYNHLVDQNFPLPD 222  
Db 181 TSSAYGGGAPLLSNTFAAGFMWLDKLGLSAQMGEVVMRQVFPFAGNHYLVDFEPLPD 240  
QY 223 YWLSLLYKRLGPKVLAVHAGLQRPGRVIRDKLRIYAHCTNHNHNHYVSGSITLFI 282  
Db 241 YWLSLLFKLVGPRVLLSRVGPD-----RSKRLRYLHCTNHYHPRYQEGDLTLIV 291  
QY 283 INLHRSRKKIKLAGTLRDKLVHQLYLPQYQEGLSKSVQLNGQPLVMVDGTLPELAPR 342  
Db 292 LNLHNVTYKHLKVPPLFRKPVDTYLLKPSGPDGLLSKSVQLNGQILKMWDEQTLFALTEK 351  
QY 343 PLRAGRTLVIPTVMGFVVKVNNALAC 370  
Db 352 PLPAGSALSPLAFSYGFFVIRDAKIAAC 379

## RESULT 7

US-09-181-336-15  
; Sequence 15, Application US/09181336A  
; Patent No. 6242238  
; GENERAL INFORMATION:  
; APPLICANT: FREEMAN, Craig Geoffrey  
; APPLICANT: PARISH, Richard  
; APPLICANT: HAMDORF, Brenton James  
; APPLICANT: HULETT, Mark Darren

; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE ENCODING MAMMALIAN  
; FILE REFERENCE: 032505-004  
; CURRENT APPLICATION NUMBER: US/09/181.336A  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER FILING DATE: 1997-10-28  
; EARLIER FILING DATE: 1997-12-09  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 15  
; LENGTH: 532  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-181-336-15

Query Match 38.1%; Score 744.5; DB 4; Length 532;

Best Local Similarity 44.5%; Pred. No. 3.9e-73;  
Matches 149; Conservative 55; Mismatches 106; Indels 25; Gaps 3;

QY 31 QFSNTYSNLI-----TEPNYRTMHGRAVNSOLGKDYIQLKSLLOPIR 75  
Db 198 QWNSNAQILLDYCSSKGYNISWELGNEPNSPLKKADIFINGSOLGEDFIQLHKLRLK-S 256  
QY 76 IYSRASLYGNIGRPNKVNIALLDGFMKVGSTVDVATWQHGYIDGRVVKVMDFLKTRLL 135  
Db 257 TFKNAKLYGPDVGQPRKTAKMLKSLFAGGGEVDSVTHWHYYLNGRTATREDFLNPDVL 316  
QY 136 DTLSDQIRKIQKVNTYTPGKKIWLEGVYVTTAGGTTNNLSYAGFLWNTLGMANOG 195  
Db 317 DIFISSVQVQFVVESTREPQKVKWLGTSAYGGGAPLLSDTFAAGFMWLDKLGLSARMG 376  
QY 196 IDVTRHSFFDHGYNHLVDQNFPLDYLWLSLLYKRLGPKVLAVHAGLQRPGRV 255  
Db 377 IEVVMRQVFFGAGNHYLVDFEPLDYLWLSLLKLVGTVKVLMAVQSGSKRR----- 429  
QY 256 RDKLRIYAHCTNHNHNHYVSGSITLFIINLHRSRKKIKLAGTLRDKLVHQLYLPQYQEG 315  
Db 430 --KLVYLHCTNTDPRYKEGDLTVYALNHNVTYRLPYFSPKQVDKYLRLPLGPHG 487  
QY 316 LKSKSVQLNGQPLVMVDGTLPELKPRLRAGRTL 350  
Db 488 LLSKSVQLNGTLTKMVDQTLPLMEKPLRPGSSL 522

## RESULT 8

US-08-435-675B-5  
; Sequence 5, Application US/08435675B  
; Patent No. 5710250  
; GENERAL INFORMATION:  
; APPLICANT: Ellis, Steven Bradley  
; APPLICANT: Williams, Mark E.  
; APPLICANT: Harpold, Michael Miller  
; APPLICANT: Schwartz, Arnold  
; APPLICANT: Brenner, Robert  
; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA: US/08/435.675B

;; FILING DATE: 05-MAY-1995  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/314,083  
;; FILING DATE: 28-SEP-1994  
;; APPLICATION NUMBER: US 07/914,231  
;; FILING DATE: 13-JUL-1992  
;; APPLICATION NUMBER: US 07/603,751  
;; FILING DATE: 08-NOV-1990  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Seidman, Stephanie L.  
;; REGISTRATION NUMBER: 33,779  
;; REFERENCE/DOCKET NUMBER: 6362-53193  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 619-238-0999  
;; TELEFAX: 619-238-0062  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1106 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; FRAGMENT TYPE: internal  
;; US-08-435-675B-5

Query Match 4.5%; Score 88; DB 1; Length 1106;  
Best Local Similarity 18.7%; Pred. No. 2;  
Matches 68; Conservative 55; Mismatches 119; Indels 122; Gaps 14;

QY 92 KNVIALDGFMKVAGSTVDVAVTWQHCHYIDGRVVKVMDFLKRLDLTLDSD----- 140  
DB 254 KDMILIVD-----VSGS-----VSGLTKLIRTSVSEMLETLSDDDDFVNVASFNS 298  
QY 141 ---QIRKIQKVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWL-----NT 187  
DB 299 NAQDVSCFQHLVQAVNRNKKVLKDAVNNITAKGITDYKKGFSFAFEQLLNNVSRANCNK 358  
QY 188 LGMLANOGIDVIRHSFFDHGYNHLV-----DQNFNPLPDYWLSL----- 227  
DB 359 IIMLFTDGEERAQEIFAKYNKDKKVRVFTFSVGQHNDRGPIQWMACENKGYEYIETSI 418  
QY 228 -----LYKRLIG-PKVLA-----VHVAGLQKRPGRVIRDKLRIYAHCTN 267  
DB 419 GAIRINTQEYLDVLRPVMVLAGDKAKQVQWTVNYLDLDEL-----GLVITGTLPVENITQ 474  
QY 268 HHNNHYVRSITLFTIINLHRSRKKIKLAGTLRDKLVHQLYLOPYQOGLKSKSVOLNGQP 327  
DB 475 FENKTNLKNQLILGVMGVDVSLEDIK-----RLTPRTLCPNGY-----YFAIDPNGYV 523  
QY 328 LVMVDDGTLPKLPRLRAGRTLVP-----PVTMGFFVVKVNNALA 369  
DB 524 LLH-----PNLQPKPIGVG-----IPTNLKRRRPNVQNPKSQEPVTLDFLDAELENDIK 573  
QY 370 CRYR 373  
DB 574 VEIR 577

RESULT 9  
US-08-336-257A-8  
; Sequence 8, Application US/08336257A  
; Patent No. 5726035  
; GENERAL INFORMATION:  
; APPLICANT: Jay, Scott D  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Harpold, Michael M.  
; APPLICANT: Campbell, Kevin P.  
; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Brown, Martin, Haller & McClain  
;; STREET: 1660 Union Street  
;; CITY: San Diego  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 92101-2926  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FastSeq version 1.5  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/336,257A  
;; FILING DATE: 07-NOV-1994  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Seidman, Stephanie L.  
;; REGISTRATION NUMBER: 33,779  
;; REFERENCE/DOCKET NUMBER: 54898  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (619) 238-0999  
;; TELEFAX: (619) 238-0062  
;; INFORMATION FOR SEQ ID NO: 8:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1106 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; FRAGMENT TYPE: internal  
;; ORIGINAL SOURCE:  
;; US-08-336-257A-8

Query Match 4.5%; Score 88; DB 1; Length 1106;  
Best Local Similarity 18.7%; Pred. No. 2;  
Matches 68; Conservative 55; Mismatches 119; Indels 122; Gaps 14;  
QY 92 KNVIALDGFMKVAGSTVDVAVTWQHCHYIDGRVVKVMDFLKRLDLTLDSD----- 140  
DB 254 KDMILIVD-----VSGS-----VSGLTKLIRTSVSEMLETLSDDDDFVNVASFNS 298  
QY 141 ---QIRKIQKVNTYTPGKKIWLEGVVTTSAGGTNNLSDSYAAGFLWL-----NT 187  
DB 299 NAQDVSCFQHLVQAVNRNKKVLKDAVNNITAKGITDYKKGFSFAFEQLLNNVSRANCNK 358  
QY 188 LGMLANOGIDVIRHSFFDHGYNHLV-----DQNFNPLPDYWLSL----- 227  
DB 359 IIMLFTDGEERAQEIFAKYNKDKKVRVFTFSVGQHNDRGPIQWMACENKGYEYIETSI 418  
QY 228 -----LYKRLIG-PKVLA-----VHVAGLQKRPGRVIRDKLRIYAHCTN 267  
DB 419 GAIRINTQEYLDVLRPVMVLAGDKAKQVQWTVNYLDLDEL-----GLVITGTLPVENITQ 474  
QY 268 HHNNHYVRSITLFTIINLHRSRKKIKLAGTLRDKLVHQLYLOPYQOGLKSKSVOLNGQP 327  
DB 475 FENKTNLKNQLILGVMGVDVSLEDIK-----RLTPRTLCPNGY-----YFAIDPNGYV 523  
QY 328 LVMVDDGTLPKLPRLRAGRTLVP-----PVTMGFFVVKVNNALA 369  
DB 524 LLH-----PNLQPKPIGVG-----IPTNLKRRRPNVQNPKSQEPVTLDFLDAELENDIK 573  
QY 370 CRYR 373  
DB 574 VEIR 577

RESULT 10  
US-08-455-543A-53  
; Sequence 53, Application US/08455543A  
; Patent No. 5792846  
; GENERAL INFORMATION:  
; APPLICANT: Harpold, Michael

APPLICANT: Ellis, Steven  
APPLICANT: Williams, Mark  
APPLICANT: Feldman, Daniel  
APPLICANT: McCue, Ann  
APPLICANT: Brenner, Robert  
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND  
TITLE OF INVENTION: METHODS  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
CITY: San Diego  
STATE: California  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,543A  
FILING DATE: May 31, 1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/223,305  
FILING DATE: April 4, 1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/868,354  
FILING DATE: April 10, 1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/745,206  
FILING DATE: 15-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/620,250  
FILING DATE: 30-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/603,751  
FILING DATE: 04-APR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US89/01408  
FILING DATE: 04-APR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/176,899  
FILING DATE: 04-APR-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-52517  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619)238-0999  
TELEFAX: (619)238-0062  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1103 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-455-543A-53

Query Match 4.4%; Score 86; DB 1; Length 1103;  
Best Local Similarity 18.7%; Pred. No. 3.3; Indels 122; Gaps 14;  
Matches 68; Conservative 54; Mismatches 120;  
Qy 92 KNVIALDGFMKVAGSTVDATWQHCYDGRVVKYMDFLKRLDLSLSD----- 140  
Db 252 KDMILVLD----VSGS-----VSGLTLKIRTSVSEMLETLSDDDFVNVSFNS 296

Qy 141 ---QIRKIQKVVNYTPGKKIWLGVVTTSGGTTNNLSDSYAAGFLWL-----NT 187  
Db 297 NAQDVSCFQHLVQANVRNKKVLDKAVNNITAKGITDYKKGFSAFQELLNYNVSRANCNK 356  
Qy 188 LGMLANQGDVVIRHSEFFDHGYNHLV-----DQNFNPLDPDYWLSL----- 227  
Db 357 IIMLETFDGEERAQEIFNKKKDKKVRFRFSVGQHNYERGPQIWMACENKGYEYIPI 416  
Qy 228 -----LYKRLIG-PKVL-----VHVAGLQKRPGRVIRDKLRIYAHCTN 267  
Db 417 GAIRINTOEYLDVLRPMVLADGKAKAQVQWTVNVLDALEL-----GLVITGTLPVENITQ 472  
Qy 268 HHNNYVRGSTITFLINLHRSKKIKLAGTLRDKLVHQLYLPQYQEGGLKSKSVOLNGQP 327  
Db 473 FENKTNLKNQLITLGMVGVDVSLDIK-----RLTPRFTLCPNGY-----YFADPNPGYV 521  
Qy 328 LVMVDDGTLPELKPRPLRAGRTLVP-----PVTMGFFVVKVNNALA 369  
Db 522 LLH-----PNLOPKPIGVG-----IPTINLKRRRNIONPKSQEPVTLDFLDAELENDIK 571  
Qy 370 CRYR 373  
Db 572 VEIR 575  
RESULT 11  
US-08-223-305C-53  
; Sequence 53, Application US/08223305C  
; Patent No. 5851824  
; GENERAL INFORMATION:  
; APPLICANT: Harpold, Michael  
; APPLICANT: Ellis, Steven  
; APPLICANT: Williams, Mark  
; APPLICANT: Feldman, Daniel  
; APPLICANT: McCue, Ann  
; APPLICANT: Brenner, Robert  
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/223,305C  
; FILING DATE: April 4, 1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/868,354  
; FILING DATE: April 10, 1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/745,206  
; FILING DATE: 15-AUG-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/620,250  
; FILING DATE: 30-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/482,384  
; FILING DATE: 20-FEB-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/603,751  
; FILING DATE: 04-APR-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US89/01408  
; FILING DATE: 04-APR-1989  
; PRIOR APPLICATION DATA:

```
; APPLICATION NUMBER: US 07/176,899
; FILING DATE: 04-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 52516 (P519739)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)238-0999
; TELEFAX: (619)238-0062
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1103 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-223-305C-53

Query Match          4.4%; Score 86; DB 2; Length 1103;
Best Local Similarity 18.7%; Pred. No. 3.3;
Matches 68; Conservative 54; Mismatches 120; Indels 122; Gaps 14;

QY  92 KNVIALLDGFMKVGSTDAVTWQHCHYIDGRVYVVMDFLTKRLDLDLSD-----140
Db  252 KDMILIVD-----VSGS-----VSGTLKLRITSVSEMLETLSDDDFNVASFNS 296
QY  141 ---QIRKIQKVNTYTPGKKIWLEGVVTTSAGTNNLSDSYAAGFLWL-----NT 187
Db  297 NQADVSCFQHLVQANVRNKKVLKDAVNNITAKGITDKGFSAPFQOLLNYNVSRANCNK 356
QY  188 LGMLANQGDVIVIRHSFFDHGYNHLV-----DQFNPLPDYWLSL-----227
Db  357 IIMLFDGGEERAQAEIFKNYKDKVRFVSFGVQHNYERGPLOWACENKGYVYIPI 416
QY  228 -----LYKRLIG-PKVLA-----VHVAGLQKRPGRVIRKDKLIYAHCTN 267
Db  417 GAIRINTQETLDVLGRPMVLGDKAKQVQWNTNVYLDALD-----GLVITGLPFFNITGQ 472
QY  268 HNNHNVRSITLFIINLHRSRKKIKLAGTLRDKLVHLYLQYGGQGLKSKSVQLNGQP 327
Db  473 FENKTLKKNOLILGVMGVDVSLIEDIK-----RLTPRTLCPNGY-----YFAIDPNGYV 521
QY  328 LVNVDDGTLPELKLPRLAGRTLIVP-----PVTMGFEVVKNNNALA 369
Db  522 LLH-----PNLQPKPIGV-----IPTINLKRPRPNIQPKSOEPVTLDFLDALENDIK 571
QY  370 CRYR 373
Db  572 VEIR 575

RESULT 12
US-08-989-925-1
; Sequence 1, Application US/08989925
; Patent No. 5989820
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN ADIPOPHILIN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

```
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/989,925
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0440 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 434 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAITUT13
; CLONE: 1620223
US-08-989-925-1

Query Match          4.3%; Score 84.5; DB 2; Length 434;
Best Local Similarity 22.1%; Pred. No. 1.1;
Matches 79; Conservative 56; Mismatches 146; Indels 77; Gaps 16;

QY  7 QHFDVMELEQREKAAQMHLLVK-EQFSNTYSNLLITEPN-----NYRTMHGRAV 55
Db  21 QQPSSV-----DRVASMPLISSTCDMVSAAYASTKESYPHVKTVCDAAEKGVRTLTA 75
QY  56 NGSQKGDYIQLKSLQPIRIYSRASLYGPNRKNVIALLDGFMKVGSTDAVTWQ 115
Db  76 SGAQ-----PILSKLEP-QIASASEYAHRLDKLEENLFILOQPTKVLADTKELVSSK 128
QY  116 HCYIDGRVVKVMDFLTKRLDLDLSDQIRKIQKVNTYTPGKKIWLEGVVTTSAGTNNLS 175
Db  129 VSGAQEMVSSAKDTVATQLSEAVDATRGAVQSGVDK-----TKSVVT---GGVQSV 177
QY  176 DSYAAGFLWNTLGMLANQGDVIVIRHSFFDHGYNHLVDQFNPLPDYWLSLLYKRLIGP 235
Db  178 GS-----RLGQMYLGSVDVTVLGKS-EEWADNHL-----PLTDAELARIATSLDGF 221
QY  236 KVLAVHVAGLQKRPGRVIR-----DKLRIYAHCTNHHNHNHNVRSITLFIINLHRSR 290
Db  222 DVASVQ-----QQREQSYFVRLGSLSERLQAHAY---EHSGLKLRATKQRAQALLQSQ 274
QY  291 KIKLAGTLR-----DKLV-----HQYLL-----QPYGQEGKSKSVQLNGQPLVMVDD 333
Db  275 ALSILMETVKGVDQKLVEGQEKLHQMWLSWNQKLOGPEKPEPPKPEQVESRALTMPRD 332

RESULT 13
US-09-175-928-4
; Sequence 4, Application US/09175928A
; Patent No. 6312921
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Mi, Sha
; APPLICANT: Genetics Institute, Inc.
```





; INFORMATION FOR SEQ ID NO: 39:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 921 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-297-494-39

Query Match	4.2%;	Score 82.5;	DB 1;	Length 921;
Best Local Similarity	24.1%;	Pred. No. 6;		
Matches 61;	Conservative 43;	Mismatches 106;	Indels 43;	Gaps 13;

  

Qy	11	VMLELQRE-KAAQMHVLVLLKE---QFS-NTYSNLIITPNNYRTMHWGRAVNGSQLG----	61
Db	228	VLYQLQETQASRCCLLVSEDNLQSCKVIQGVLEEEISPLATGR-----LGQVVE	281
Qy	62	-KDYIQLKSL---LQPIR---IYSRASLYGPNIGRPKNVIALLDGFMKVAG---STV	109
Db	282	DKKSIQLKDLTSEDQQQLQSMIGCEVQAMLCVPVISRATDQVALACAFNKLGGDLFTDQ	341
Qy	110	DAVTWQHC--YIDGRVVKVMDFLKTRLLDLSQIRKIKQKVNTYTPGKKIWLEGVWTTTS	167
Db	342	DEHVIQHCFTYTSVLTSTLAFQKEQKLCQECQALLQVAKNLFTHLDDVSVLLQELITEA	401
Qy	168	AGGTNNLSDSYAAGFLWL--NTL-----GMLANQIGIDVVIHRSFFDHGYNHLVDQNFN	218
Db	402	----RNLSNAEICSVFLLDQNELVAKVFDGGVVEDESEVIRIPADQGIAGHVATTGQILN	457
Qy	219	PLPDYWLSLLYKR	231
Db	458	-IPDAYAHPLEYR	469

Search completed: July 30, 2002, 08:15:09  
Job time: 291 sec

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**